

TRANSKRIPTOM DES MERISTEMGEWEBES DES KOLUMNARAPFELS

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

(Dr. rer. nat.)

am Fachbereich Biologie der
Johannes Gutenberg-Universität in Mainz



Clemens Krost

Geboren am 12. März 1985 in Mainz

Im Februar 2012

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 27.03.2012

Für meine wundervolle Frau

*Des Winters letzte Eiseskälte,
Kündigte Frühling an in Bälde.*

*Sobald der Reif Vergangenheit,
Machte der Baum sich wuchsbereit.*

*Die erste Wärme lockte dann,
Blätter und Triebe aus dem Stamm.*

*Die Blüte kam, die Blüte ging,
Am Stamme da ein Apfel hing.*

*Viel sonn'ge Tage, etwas Regen,
Machten den Apfel ganz verlegen.*

*Rotgolden, knackig, wunderbar,
Schmeckt er dem jungen Liebespaar.*

*Entspannt, verträumt und sehr verliebt,
Es in des Baumes Schatten liegt.*

*Wird wieder eisig dann die Luft,
kommt aus dem Haus Bratapfelduft.*

*Im Garten einsam steht er da,
Der Baum sich freut auf nächstes Jahr.*

Clemens Krost

INHALTSVERZEICHNIS

1	Einleitung.....	1
1.1	Der Kulturapfel, <i>Malus x domestica</i>	1
1.2	Regulation und Kontrolle des Wachstums bei Gefäßpflanzen	3
1.2.1	Auxin.....	4
1.2.2	Cytokinin	6
1.2.3	Gibberellin	8
1.2.4	Brassinosteroid.....	9
1.2.5	Jasmonsäure (Jasmonat)	11
1.2.6	Abscisinsäure	13
1.3	Kolumnarwachstum beim Apfel.....	14
1.4	Genexpressionsanalyse mit Hilfe von Next Generation Sequencing (NGS)	18
1.5	Zielsetzung	20
2	Material und Methoden	23
2.1	Pflanzenmaterial	23
2.2	Molekularbiologische Methoden.....	24
2.2.1	RNA-Extraktion.....	24
2.2.2	DNase-Verdau	24
2.2.3	Ethanol-fällung.....	25
2.2.4	Poly-A ⁺ -Aufreinigung.....	25
2.2.5	RNA-Amplifikation	26
2.2.6	Agarose-Gelelektrophorese	26
2.2.7	Qualitäts- und Konzentrationsbestimmung der RNA.....	27
2.2.8	cDNA-Synthese	27
2.2.8.1	Sequenzierungen.....	27
2.2.8.2	qRT-PCR.....	28
2.2.9	PCR und Aufreinigung.....	28

2.2.10	Klonierung von PCR-Fragmenten.....	28
2.2.11	Plasmid-Präparation.....	29
2.2.12	Sanger-Sequenzierung	29
2.2.13	Microarray-Analysen	29
2.2.14	Quantitative Real-Time-PCR.....	29
2.2.15	Keimung von Apfelkernen.....	30
2.2.16	Protoplastenisolierung	30
2.2.17	Chromosomenpräparate	31
2.2.18	Fluoreszenz <i>in situ</i> -Hybridisierung (FISH)	31
2.2.18.1	Vorbereitung der Präparate	31
2.2.18.2	Sonden.....	32
2.2.18.3	Hybridisierung und Detektion	32
2.2.19	Puffer und Lösungen.....	32
2.2.20	Primer	34
2.3	Next Generation Sequencing.....	37
2.3.1	Roche 454	37
2.3.2	Illumina Genome Analyzer IIx	37
2.3.3	Illumina HiSeq 2000	37
2.4	Bioinformatische Methoden	38
2.4.1	Konvertierung der Datenformate.....	38
2.4.2	Analyse der Rohdaten.....	39
2.4.3	Assemblierung der Rohdaten	39
2.4.3.1	Assemblierung der Roche 454-Rohdaten.....	39
2.4.3.2	Assemblierung der Illumina-Rohdaten.....	39
2.4.3.3	Assemblierung kleinerer Datenmengen	40
2.4.4	BLAST.....	40
2.4.5	Verarbeitung der BLAST-Ausgabe mit BORT.....	41
2.4.6	Statistische Analyse der differentiellen Regulation	41
2.4.6.1	RPKM-Normalisierung.....	42
2.4.7	Kartierung von Rohdaten	42
2.4.7.1	PASS.....	42
2.4.7.2	Bowtie.....	43
2.4.8	Funktionelle Analyse mit MapMan	44
2.4.9	Analyse der differentiellen Expression mit DESeq.....	44

3	Ergebnisse	45
3.1	Extraktion und Aufarbeitung der RNA.....	45
3.2	Sequenzierungen und Analyse der Rohdaten.....	47
3.3	Assemblierung	53
3.4	Verarbeitung der Sequenzinformationen.....	57
3.4.1	Datenpipeline.....	57
3.4.2	Differentielle Regulation	62
3.5	Transkriptomdaten und Kolumnarwachstum	65
3.5.1	Globaler Vergleich der Datensätze	65
3.5.2	Komparative Genexpressionsstudien.....	70
3.5.2.1	Microarrays.....	74
3.5.2.2	qRT-PCRs	76
3.5.3	Hormonhaushalt im kolumnaren Sprossapikalmeristem	77
3.5.4	Lokalisation differentiell regulierter Gene.....	79
3.6	Chromosomenpräparate und FISH.....	82
4	Diskussion.....	85
4.1	Einfluss von Materialentnahme und -verarbeitung	85
4.2	NGS und Datenverarbeitung.....	86
4.3	Hohe Anzahl differentiell regulierter Gene im SAM kolumnarer Bäume ...	90
4.4	Differentiell regulierte Gene und kolumnarer Phänotyp	91
4.5	Phytohormonhaushalt des kolumnaren SAMs	93
4.6	Co-Allel und Beeinflussung benachbarter Gene.....	99
5	Zusammenfassung.....	103
	Abkürzungsverzeichnis	104
	Abbildungsverzeichnis	107
	Tabellenverzeichnis	108
	Literaturverzeichnis.....	109
	Supporting Information [SI]	126
	Curriculum Vitae	130
	Publikationsliste	132
	Danksagung.....	133
	Eidesstattliche Erklärung.....	135

1 EINLEITUNG

1.1 DER KULTURAPFEL, *MALUS X DOMESTICA*

Äpfel sind die meistverbreiteten Früchte der gemäßigten Zonen und werden seit der Antike in Europa und Asien kultiviert [1], weshalb der Apfel einen großen Stellenwert in unserer Symbolik und Kultur einnimmt. So ist er als Liebesapfel ein uraltes Symbol für Liebe, Sexualität und Fruchtbarkeit und wird u. a. im Hohelied Salomons und Goethes Faust in dieser Hinsicht verwendet. Aus Legenden des Apfelbaums als Baum des ewigen Lebens entstanden Aphorismen wie „An apple a day keeps the doctor away“. Der unerlaubte Genuss des Apfels vom Baum der Erkenntnis im christlichen und jüdischen Glauben führte zum Sündenfall im Paradies (1. Buch Mose). Der Reichsapfel war jahrhundertlang ein Zeichen des Besitzanspruches im mitteleuropäischen Kaisertum. Als Zankapfel diente er als Werkzeug der Göttin Eris (Discordia), um die Damen der Hochzeitsgesellschaft von Peleus und Thetis in heftigen Streit zu versetzen.

Die Gattung *Malus* hat etwa 25-30 Spezies [2] und einige Subspezies der sogenannten Holz- oder Johannesäpfel, die aufgrund ihrer attraktiven Blüten und des Blattwerkes als Zierbäume kultiviert werden. Der heutige Kulturapfel resultiert sehr wahrscheinlich aus einer interspezifischen Hybridisierung, daher ist die Namensgebung *Malus x domestica* Borkh. als wissenschaftlicher Name akzeptiert [3]. Der Kulturapfel geht auf den Vorgänger *Malus sieversii* zurück, dessen natürliches Vorkommen sich vom Tien Shan Gebirge an der Grenze von China und Kirgisien/Kasachstan bis hin zum kaspischen Meer erstreckt [4, 5], weshalb Zentralasien (besonders um Alma Ata) mit seiner großen Diversität als Ursprung („Vater“) des Apfels gilt [6]. Seit 2006 werden Wildapfel und domestizierter Apfel unter dem Namen *Malus pumila* vereinigt, da es sich um die gleiche Art handelt [7]. Auch wenn diese Erkenntnis lange Zeit in Frage gestellt wurde [8], gilt die Hypothese spätestens seit der Sequenzierung des Genoms als bewiesen [9]. Dennoch wird in dieser Arbeit die Bezeichnung *Malus x domestica* beibehalten.

Der Apfel ist aufgrund seiner hohen Lagerfähigkeit und seiner multiplen Verarbeitungsmöglichkeiten sehr populär; so gehören Tafeläpfel, Saft, Kompott, Apfelwein und Cidre zu den häufigsten Apfelprodukten. Des Weiteren gilt er u. a. aufgrund seiner krebsvorsorgenden Wirkung [10] seit langem als Heilmittel. Mit einer Produktion von über 69 Millionen Tonnen jährlich [11] rangiert der Apfel unter den wirtschaftlich bedeutendsten Obstsorten auf Platz vier [12]. Die am weitesten verbreiteten Sorten sind ‚Delicious‘ und ‚Golden Delicious‘ [1]. Die ersten kontrollierten Kreuzungen gehen zurück auf Thomas A. Knight im Jahre 1806. Heutige „Hochgeschwindigkeitskreuzungen“ und gentechnologische Ansätze zielen meist auf die Weiterentwicklung zu krankheitsresistenten, hochqualitativen Sorten ab [13, 14].

Botanisch gesehen gehört der Apfel zur Familie der *Rosaceae* (Rosengewächse), Unterfamilie *Pomoideae* (Kernobst). Weitere Vertreter dieser Familie sind u.a. Birne, Quitte und Mispel. Im Gegensatz zu anderen Unterfamilien der *Rosaceae* (in der Regel 7-9 Chromosomen) besitzen die *Pomoideae* 17 Chromosomen [15] und sind in der Regel diploid ($2n = 34$), wobei auch tri- und tetraploide Sorten existieren [16]. Aufgrund der stark ausgeprägten Heterozygotie und Selbstinkompatibilität sind Kreuzungen im Hinblick auf einzelne, gewünschte Faktoren äußerst schwierig. Daher geschieht die Vermehrung der Kultursorten meist klonal oder vegetativ durch Pfropfen auf eine geeignete Unterlage. Hauptproblematik in der herkömmlichen Züchtung ist die lange juvenile Phase, in der keine Fruchtbildung stattfindet. Hier könnten neueste Erfolge in der Hochgeschwindigkeitskreuzung [13] und Transformation [17-19] Abhilfe schaffen und richtungsweisend für die Etablierung neuer Sorten sein.

Voraussetzung für züchterische Erfolge ist die genaue Kenntnis des Genoms. Dieses wurde im Jahre 2010 anhand der Sorte ‚Golden Delicious‘ entschlüsselt [9]. Danach besitzt der Apfel einen haploiden Chromosomensatz von $n = 17$ mit einer Gesamtgenomgröße von 742,3 Mb (Megabasen). Davon sind 67 % repetitive Elemente, 38 % entfallen alleine auf Retrotransposons. Die Arbeitsgruppe um Riccardo Velasco konnte eine starke Kolinearität der Chromosomen 3 und 11, 5 und 10, 9 und 17, 13 und 16 sowie weiterer kleiner Abschnitte zeigen, was auf eine genomweite Duplikation (GWD) vor 30-45 Millionen Jahren, vermutlich durch Autopolyploidisierung eines Vorfahren mit neun Chromosomen ($n = 9$) bei Verlust eines Chromosoms, zurückzuführen ist. Dadurch ergibt sich die ungewöhnlich hohe Zahl vorausgesagter Gene von 57,386 (zum Vergleich: Arabidopsis 26,500; Reis

41,000; Pappel 40,000 [20]; Erdbeere 34,108 [21]; Kakao 27,798 [22], Soja 46,430 [23]). Darunter findet man ungewöhnlich viele Transkriptionsfaktoren und MADS-Box-Gene sowie 11,444 apfelspezifische Gene.

1.2 REGULATION UND KONTROLLE DES WACHSTUMS BEI GEFÄßPFLANZEN

Als Wachstum wird eine irreversible Volumenzunahme einer Zelle, eines Gewebes oder Organs bezeichnet. Differenzierung ist dagegen eine qualitative Veränderung der Form oder Funktion [24]. Im Unterschied zu Tieren sind höhere Pflanzen aus Phytomeren aufgebaut und somit modular. Diese Phytomere werden von den Scheitelmeristemen, also Spross- und Wurzelmeristem, gebildet. Schon während der Embryogenese werden diese primären Meristeme angelegt, aus denen alle weiteren Organe der Pflanze hervorgehen. Die Zellen der Meristeme sind damit pluripotent, teilweise sogar totipotent. Das Zellwachstum beinhaltet die Vermehrung von Zellbestandteilen (Plasmawachstum) und die Vergrößerung des Zellvolumens (Streckungswachstum), wobei letztere gleichmäßig (isodiametrisch) oder gerichtet (prosenchymatisch) vonstattengehen kann. Das Streckungswachstum ist stets mit der Vergrößerung der Vakuolen und Bildung einer Zentralvakuole durch Wassereinlagerung verknüpft. Der Anteil von löslichen Zellbestandteilen und Zellwandmaterial nimmt während dieses Prozesses, wenn überhaupt, nur mäßig zu. Die Zellstreckung findet in den Streckungszonen nahe der Meristeme statt, die Zellvermehrung (Teilungswachstum) dagegen ausschließlich im Meristemgewebe selbst (Längenwachstum in den Apikal- und sekundären Blattachselmeristemen, Dickenwachstum im Kambium).

Das Pflanzenwachstum unterliegt komplexen Regulationsmechanismen, die auf die sessile Lebensweise abgestimmt sind. So beeinflussen endogene Signale (Alter, Entwicklungsphase, vegetative oder generative Ausrichtung) und Außenfaktoren (Temperatur, Licht, Nährstoffangebot, Feuchtigkeit, Gravitation) das momentane Wachstum [25]. Auch auf Stresssituation abiotischer (Trockenheit, Salz, Kälte, Hitze) und biotischer Natur (Insektenfraß, Pathogenbefall) müssen Pflanzen reagieren können. Diese Stimuli werden auf verschiedene Weisen erkannt und in chemische Botenstoffe umgewandelt, die als Hormone innerhalb der Pflanze, als Pheromone zwischen Pflanzen der gleichen Art, oder als Kairomone wirken [26], die

beispielweise die Anwesenheit von bestimmten Fressfeinden vortäuschen. Den höchsten Stellenwert nehmen aber die Phytohormone ein, endogene niedermolekulare Substanzen, die in kleinsten Mengen (ca. 10^{-6} bis 10^{-8} M) an ihrem Bildungsort oder nach Transport durch Xylem oder Phloem an ihrem Bestimmungsort wirken und dort das Wachstum beeinflussen. Die Signaltransduktion erfolgt meist über Rezeptorfamilien, welche G-Proteine, Inositolphosphate, Proteinkinasen oder die Ubiquitin-abhängige Proteindegradation aktivieren oder inhibieren, was letztendlich in geänderter Genexpression im Zielgewebe resultiert. Oft haben Phytohormone multiple Effekte, die sich im Hinblick auf das Wachstum durchaus überschneiden können [27] und sowohl synergistisch wie auch antagonistisch wirken [28]. Weiterhin beeinflussen sich Phytohormone in ihrer Wirkung und Regulation untereinander und bilden dadurch komplexe Interaktionsnetzwerke, die noch weitgehend unverstanden sind [29, 30]. In den folgenden Abschnitten werden die bekanntesten Vertreter der Phytohormone vorgestellt, bei denen man aufgrund von Mutationen in einzelnen Komponenten der Biosynthese und Signaltransduktion phänotypische Veränderungen identifiziert hat, die sich auch mit dem Kolumnarwachstum assoziieren lassen.

1.2.1 AUXIN

Auxine (von griechisch „auxein“ – wachsen) sind eine Gruppe natürlicher und synthetischer Phytohormone, deren strukturelle Gemeinsamkeit in einem aromatischen, planaren Ringsystem und einem hydrophoben Linker mit terminaler Carboxylgruppe besteht. Neben Cytokininen sind Auxine die einzige Gruppe von Phytohormonen, die für das Überleben der Pflanze essentiell sind. Mit ihren pleiotropen Effekten sind Auxine an nahezu allen Entwicklungsprozessen in der Pflanze beteiligt, wobei sie sowohl Teilungs- und Streckenwachstum wie auch Differenzierungsprozesse beeinflussen. Das am häufigsten natürlich vorkommende Auxin ist Indol-3-essigsäure (indole-3-acetic acid, IAA), die vornehmlich in jungen Blättern und im Sprossapex synthetisiert wird. IAA ist aufgrund der dosisabhängigen Wirkung auf die Entwicklung ein pflanzliches Morphogen [31].

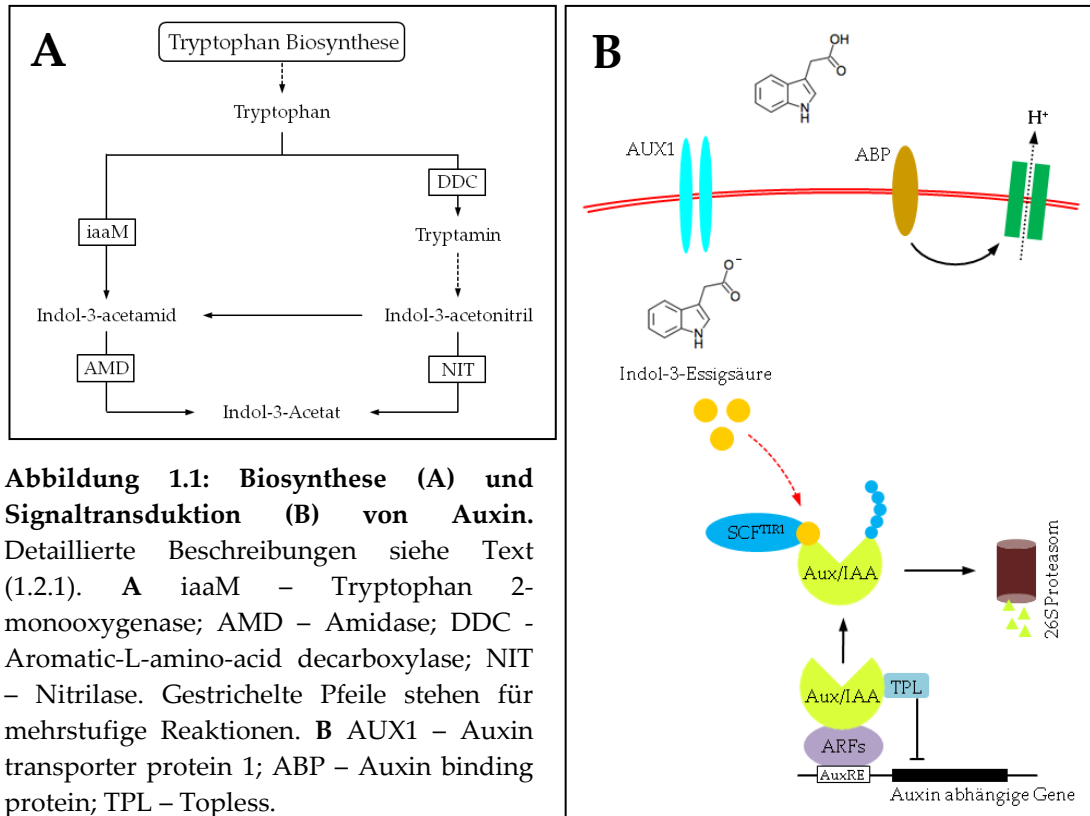


Abbildung 1.1: Biosynthese (A) und Signaltransduktion (B) von Auxin. Detaillierte Beschreibungen siehe Text (1.2.1). **A** *iaaM* – Tryptophan 2-monooxygenase; *AMD* – Amidase; *DDC* – Aromatic-L-amino-acid decarboxylase; *NIT* – Nitrilase. Gestrichelte Pfeile stehen für mehrstufige Reaktionen. **B** *AUX1* – Auxin transporter protein 1; *ABP* – Auxin binding protein; *TPL* – Topless.

Ihre Biosynthese ist sehr komplex und kann sowohl Tryptophan (Trp)-abhängig wie Trp-unabhängig ablaufen, wobei noch nicht final geklärt ist, welche Synthesewege tatsächlich von Pflanzen benutzt werden [32]. Bestätigt scheinen jedoch die Trp-abhängige Synthese über Indol-3-acetamid und Indol-3-acetonitril (Abbildung 1.1 A) [33]. Darüber hinaus gewinnen Pflanzen auch über die hydrolytische Spaltung von inaktiven IAA-Konjugaten wie IAA-Aminosäuren (AS), IAA-Zucker und IAA-Methylester freie, aktive IAA [34, 35]. Der Transport findet entweder über den Massenstrom oder polar von Zelle zu Zelle statt. Letzterer läuft unidirektional-basipetal ab und erzeugt so einen Auxingradienten im Pflanzenkörper. Gradient und Transport sind entscheidend für Auxin-abhängige Mechanismen wie Tropismen [36], bei Apikaldominanz und apikaler Kontrolle mit Cytokinin als Antagonist [37, 38] und im Zusammenspiel mit Cytokinin und Strigolaktone bei der Kontrolle der Verzweigung [39]. Der polare Transport wird über Auxinflux (LAX) und -efflux (PIN) Proteine reguliert, die asymmetrisch im Plasmalemma verteilt vorliegen [40, 41]. Nicht abschließend geklärt ist die Frage nach dem/den Rezeptor/en. Zum einen existieren Auxin-bindende Proteine (ABP) [42], deren Aktivierung zur Erhöhung der H⁺-ATPase-Aktivität des Plasmalemmas und damit zur Ansäuerung der Zellwand (Wachstum) führt (Abbildung 1.1 B) [43], hierbei

wird ebenfalls die mögliche Aktivität eines G-Proteins in der Signaltransduktion diskutiert [44]. Andererseits bedingt die Anwesenheit des Hormons die spezifische Aktivierung Auxin-abhängiger Gene der drei Genfamilien GH3, SAUR und Aux/IAA, wobei eine andere Signaltransduktionskaskade zu Grunde liegt (Abbildung 1.1 B) [27]. Intrazelluläre IAA bindet an den Rezeptor Transport Inhibitor Response (TIR1), ein F-Box-Protein, welches mit Cullin (CUL) und S-phase kinase-associated protein 1 (SKP1) als SCF^{TIR1}-E3-Ubiquitin-Ligase-Komplex fungiert. Durch diese Bindung werden Aux/IAA-Proteine ubiquitiniert und im 26S-Proteasom degradiert. Dadurch werden Auxin Response Factors (ARFs) aktiv, die normalerweise durch Dimerisierung mit Aux/IAA-Proteinen inhibiert sind, wozu ein weiterer Faktor TOPLESS (TPL) notwendig ist [45]. Die frei werdenden ARFs binden an Auxin Response Elements (AuxREs) und aktivieren auf diese Weise die oben genannten Gene.

1.2.2 CYTOKININ

Cytokinine sind N⁶-substituierte Purine, die ihren Namen der Eigenschaft verdanken, fördernd auf das Teilungswachstum von Zellen (Cytokinese = Zellteilung) zu wirken. Sie werden durch Übertragung von DMAPP (Dimethylallylpyrophosphat) auf ATP, ADP oder AMP gebildet (Abbildung 1.2 A), gefolgt von weiteren Umsetzungen durch Cytokininhydroxylasen (CYP735A1/2) [46]. Zu den häufigsten Vertretern gehören N⁶-Isopentenyladenin (IPA) und trans-Zeatin (tZ), die sowohl als freie Basen, Riboside wie auch Ribotide vorliegen, wobei nur erstere physiologisch aktiv sind [47-49]. Die Bildung von Cytokininen geschieht in teilungsaktiven Geweben, insbesondere Wurzelspitzen, der Transport erfolgt über das Xylem [50]. Bioaktive Cytokinine können durch Glucosyltransferasen reversibel O-glucosyliert oder irreversibel N-glucosyliert und damit inaktiviert werden [50, 51].

Neben der bereits angesprochenen Wirkung auf die Zellteilung sind Cytokinine auch an Blattseneszenz, Chloroplastenentwicklung, Gefäßdifferenzierung und Knöllchenbildung beteiligt [52, 53]. Des Weiteren spielen sie als Gegenspieler zu Auxin eine wichtige Rolle in der Aufrechterhaltung der Apikaldominanz, denn Auxin hemmt die Akkumulation von Cytokinin in den Seitenknospen und damit deren Austreiben [54].

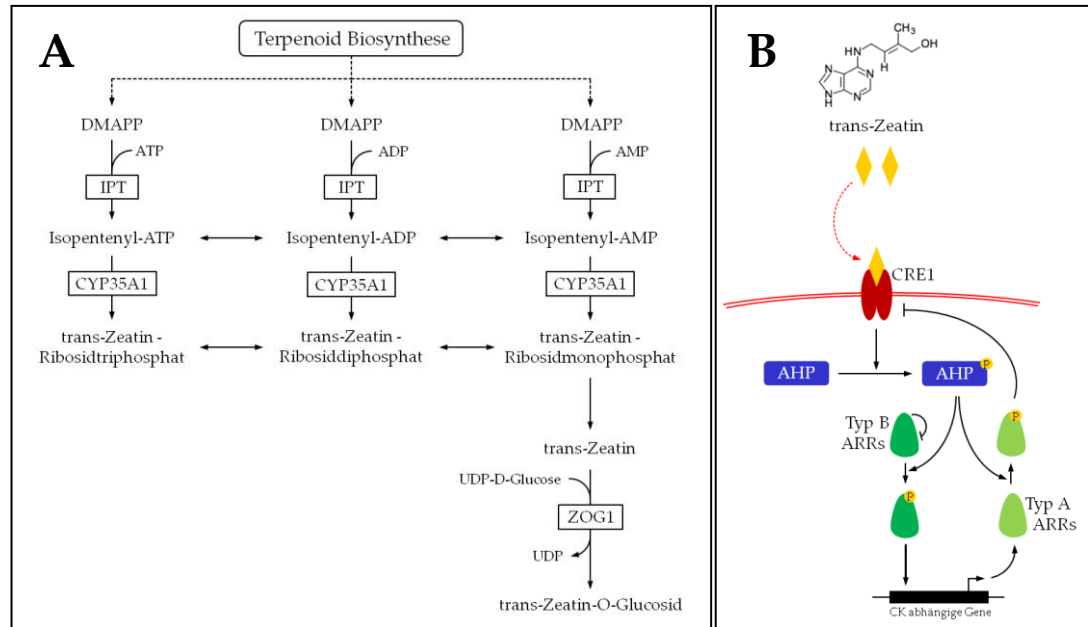


Abbildung 1.2: Biosynthese (A) und Signaltransduktion (B) von Cytokinen (CK) am Beispiel von trans-Zeatin (tZ). Detaillierte Beschreibungen siehe Text (1.2.2). **A** IPT – Isopentenyltransferase; CYP735A1 – Cytokinin hydroxylase A1; ZOG1 – Zeatin O-Glucosyltransferase 1. Gestrichelte Pfeile stehen für mehrstufige Reaktionen. **B** CRE1 – Cytokinin-Resistant 1; AHP – *Arabidopsis* Histidine Phosphotransferase; ARR – *Arabidopsis* Response Regulator.

Die Signaltransduktionskaskade von Cytokinin (Abbildung 1.2 **B**) verläuft über ein Zweikomponenten-Phosphotransfer-System [55]. CRE1 (Cytokinin-RESistent) und AHKs (Histidine Kinases) bilden dabei hybride Rezeptordimere, die durch Bindung von Cytokinin an einem konservierten Histidinrest autophosphoryliert werden [56]. Dieser Phosphatrest wird dann auf einen konservierten Aspartatrest im AHK übertragen und anschließend auf ein AHP (*Arabidopsis* Histidine Phosphotransferase) transferiert. Das AHP wandert aktiv in den Zellkern und phosphoryliert dort Typ-A- oder Typ-B-ARRs (*Arabidopsis* Response Regulators). Aktivierte Typ-B-ARRs steuern die Expression Cytokinin-abhängiger Gene, darunter auch Typ-A-ARRs. Diese dienen in der aktiven Form als negative Regulatoren der Cytokinin-Signaltransduktion [57].

bildet. Anschließend erfolgt eine Hydroxylierung über Gibberellin-3-beta-dioxygenase (GA3ox). Alternative Umsetzungen über P450-3 (Cytochrom-P450-Monooxygenase 3) und DES (GA₄ Desaturase) wurden bislang nur in Pilzen, nicht aber in Pflanzen nachgewiesen [59, 60]. Neben der Aktivität von GA3ox ist überdies GA2ox (Gibberellin-2-beta-dioxygenase) an der Regulation der Menge von aktivem Hormon in der Zelle entscheidend beteiligt, da GA2ox aktive GAs durch Oxidation konstitutiv inaktiviert [61]. Reversible Inaktivierung erfolgt dagegen durch Konjugation mit Zuckern. Gibberelline konnten zwar im Phloem- und Xylemsaft identifiziert werden, was ihren Transport beweist, die Synthese findet aber überwiegend direkt am Wirkungsort statt.

Ähnlich wie bei Auxin und BRs wird die Zellstreckung durch eine Veränderung der Zellwand induziert, an der Xyloglucan Endotransglucosylasen/hydrolasen (XTHs) und Expansine beteiligt sind [62]. Die vorgeschaltete Signaltransduktion (Abbildung 1.3 B) verläuft ähnlich wie bei Auxin: GA bindet an seinen cytoplasmatischen Rezeptor GID1 (GA-Insensitive Dwarf 1), der mit den transkriptionsreprimierenden DELLA-Proteinen interagiert. Diese Interaktion führt zur Rekrutierung des SCF^{GID2/SLY1}-Komplexes, woraufhin die DELLA-Proteine ubiquitiniert und proteasomal abgebaut werden. Zuvor inhibierte Transkriptionsfaktoren (Phytochrome Interacting Factors, PIFs) sorgen so für die Expression GA-abhängiger Gene.

1.2.4 BRASSINOSTEROID

Brassinosteroide (BRs) mit ihrem wirksamsten Vertreter Brassinolid (BL) sind eine Gruppe von Steroidhormonen, die verschiedene Wachstums- und Entwicklungsprozesse, darunter Zellteilung und Zellstreckung, regulieren [63]. Letztere erfolgt ähnlich wie bei Auxin durch Ansäuerung des Apoplasten und Induktion von Expansinen und XTHs. Da dies allerdings verzögert geschieht, basiert die Reaktion wahrscheinlich auf einer Änderung der Genexpression [64]. Wie Auxin sind auch BRs an der Aufrechterhaltung der Apikaldominanz beteiligt. BRs werden vermutlich nicht oder nur in geringen Maßen transportiert, sondern direkt an ihrem Wirkungsort synthetisiert [65].

Die Biosynthese (Abbildung 1.4 A) zweigt wie bei CK, ABA und GA von der Terpenoid-Biosynthese ab, wobei Squalen bzw. Campesterol die

Ausgangssubstanzen darstellen und deren Synthese damit der geschwindigkeitsbestimmende Schritt ist. Die C₂₇-, C₂₈-, und C₂₉-Steroide werden daraus unter Beteiligung verschiedener Cytochrom P450 Monooxygenasen (CYP) gebildet [66]. BRs werden von einer membranständigen, rezeptorähnlichen Kinase BRI1 (BR Insensitive 1) perzipiert (Abbildung 1.4 B), die bei Abwesenheit von BRs von BKI1 (BRI1 Kinase Inhibitor 1) inhibiert wird.

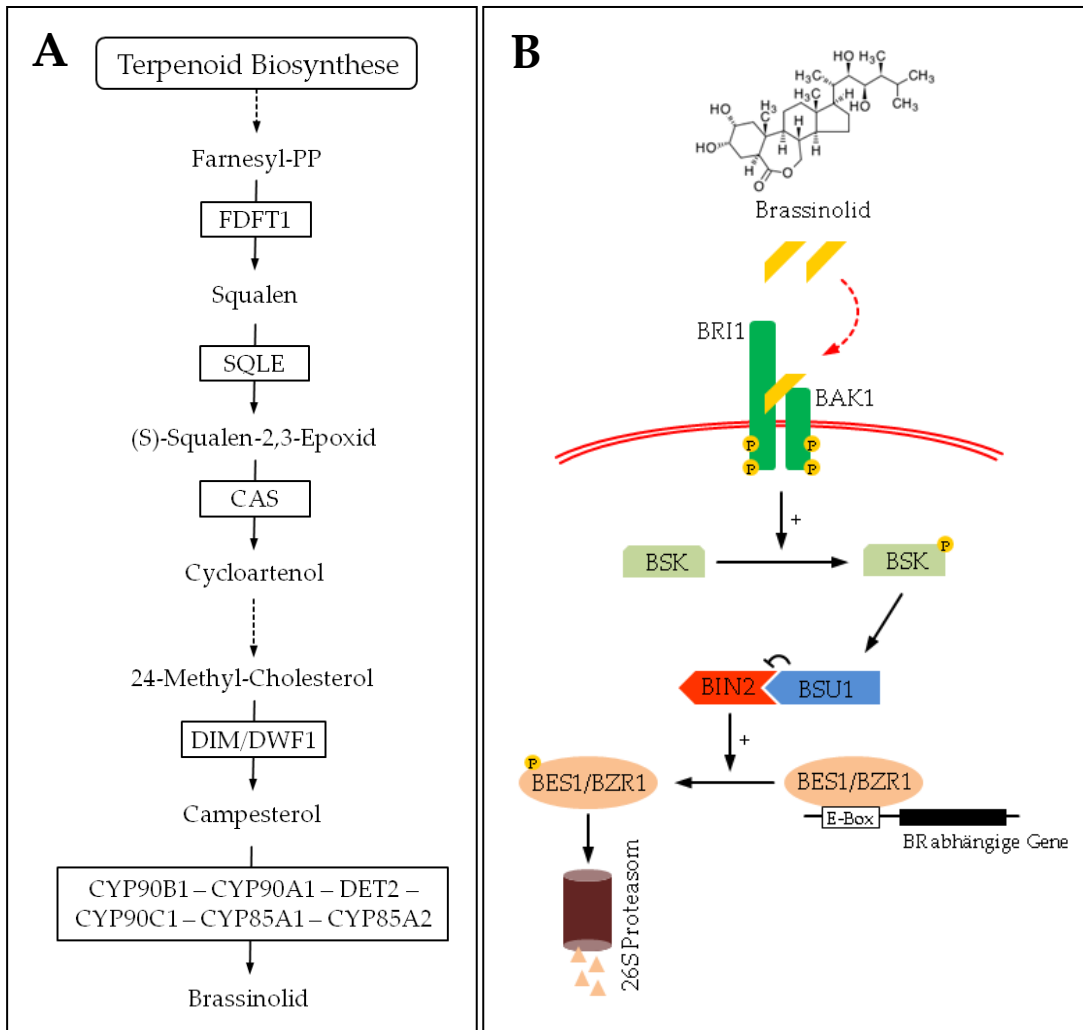


Abbildung 1.4: Biosynthese (A) und Signaltransduktion (B) von Brassinosteroiden (BRs) am Beispiel von Brassinolid (BL). Detaillierte Beschreibungen befinden sich im Text (1.2.4). **A** FDT1 – Farnesyl-Diphosphate Farnesyltransferase 1; SQLE – Squalene Monooxygenase; CAS – Cycloartenol Synthase; DIM/DWF1 - Diminuto/Dwarf1; DET2 - Probable Steroid Reductase DET2. Gestrichelte Pfeile stehen für mehrstufige Reaktionen. **B** BRI1 – BR Insensitive 1; BAK1 – BRI1-Associated Receptor Kinase 1; BSK – BR Signaling Kinase; BSU1 – BR Suppressor 1; BIN2 – BR Insensitive 2; BES1/BZR1 – BRI1-EMS Suppressor 1/Brassinazole Resistant 1.

Die Präsenz von BRs führt dagegen zur Phosphorylierung des Corezeptors BAK1 (BRI1-Associated Receptor Kinase 1), der wiederum BRI1 phosphoryliert und damit aktiviert. Aktives BRI1 phosphoryliert daraufhin seine Substrate, darunter BSK (BR Signaling Kinase), das so mit der Phosphatase BSU1 (BR Suppressor 1) assoziieren kann. BSU1 inhibiert durch diese Assoziation die GSK3/Shaggy-like Kinase BIN2 (BR Insensitive 2), die in Abwesenheit von BRs bHLH-ähnliche Transkriptionsfaktoren der BES1 (BRI1-EMS Suppressor 1)- und BZR1 (Brassinazole Resistant 1)-Familie phosphoryliert. Dadurch werden diese proteasomal abgebaut oder durch 14-3-3-Proteine aus dem Zellkern transportiert [67, 68]. Dephosphoryliert (in Anwesenheit von BRs) sind BES1 und BZR1 aktiv, wandern in den Zellkern, interagieren mit weiteren TFs und induzieren dadurch die Expression BR-abhängiger Gene [69].

1.2.5 JASMONSÄURE (JASMONAT)

Jasmonate (engl. Jasmonic Acid, JA) stellen eine Gruppe fettsäurebasierter Hormone (Oxylipide) dar, die eine hohe Ähnlichkeit zu tierischen Prostaglandinen aufweisen. JAs spielen eine wesentliche Rolle bei der Antwort auf biotischen und abiotischen Stress [70] und entscheiden durch transkriptionelle Umprogrammierung der Zelle, ob diese weiter wächst oder sich verteidigt (Wachstumshemmung) [71]. Dies wird dadurch realisiert, dass abhängig vom jeweils vorherrschenden, aktiven Hormon unterschiedliche Subsets von Genen exprimiert werden [72].

Die Biosynthese von JA (Abbildung 1.5 A) geht von α -Linolensäure aus und beinhaltet eine Serie von Reaktionen, die Zyklisierungs-, Reduktions- und Oxidationsschritte beinhalten. Diese werden in Chloroplasten und Peroxisomen von den Enzymen Lipxygenase (LOX), Allene Oxide Synthase (AOS), Allene Oxide Cyclase (AOC) und 12-Oxophytodienoate Reductase (OPR) katalysiert [73]. Das Endprodukt (+)-7-Iso-JA epimerisiert in die stabilere (-)-JA und wird von JAR1 (Jasmonate Resistant 1) mit Aminosäuren konjugiert und somit aktiviert. Viele JA-Konjugate zeigen keine hormonale Wirkung; als die biologisch wirkungsvollste Variante gilt (-)-JA-L-Ile [74].

Die Signaltransduktion (Abbildung 1.5 B) ist wie bei Auxin und GA mit der proteasomalen Degradation von Transkriptionsrepressoren gekoppelt. In diesem Fall sind dies die Jasmonate ZIM-Domain (JAZ)-Proteine [75-77]. Durch die

Bindung von aktivem Hormon kommt es zur Bildung von JAZ-CO1 (Coronatine Insensitive 1) Komplexen. CO1 ist das F-Box-Protein des SCF^{CO1}-Komplexes, welcher die Ubiquitinierung der JAZ-Proteine bewirkt und damit deren Degradation am 26S-Proteasom einleitet. In Abwesenheit von aktiver JA reprimieren die JAZ-Proteine TFs wie MYC2 durch Interaktionen mit NINJA (Novel Interactor of JAZ) und dem Corepressor TPL (TOPLESS). Diese Art der Interaktion lässt sich ebenfalls bei der Auxin- und ABA-Signaltransduktion nachweisen [78].

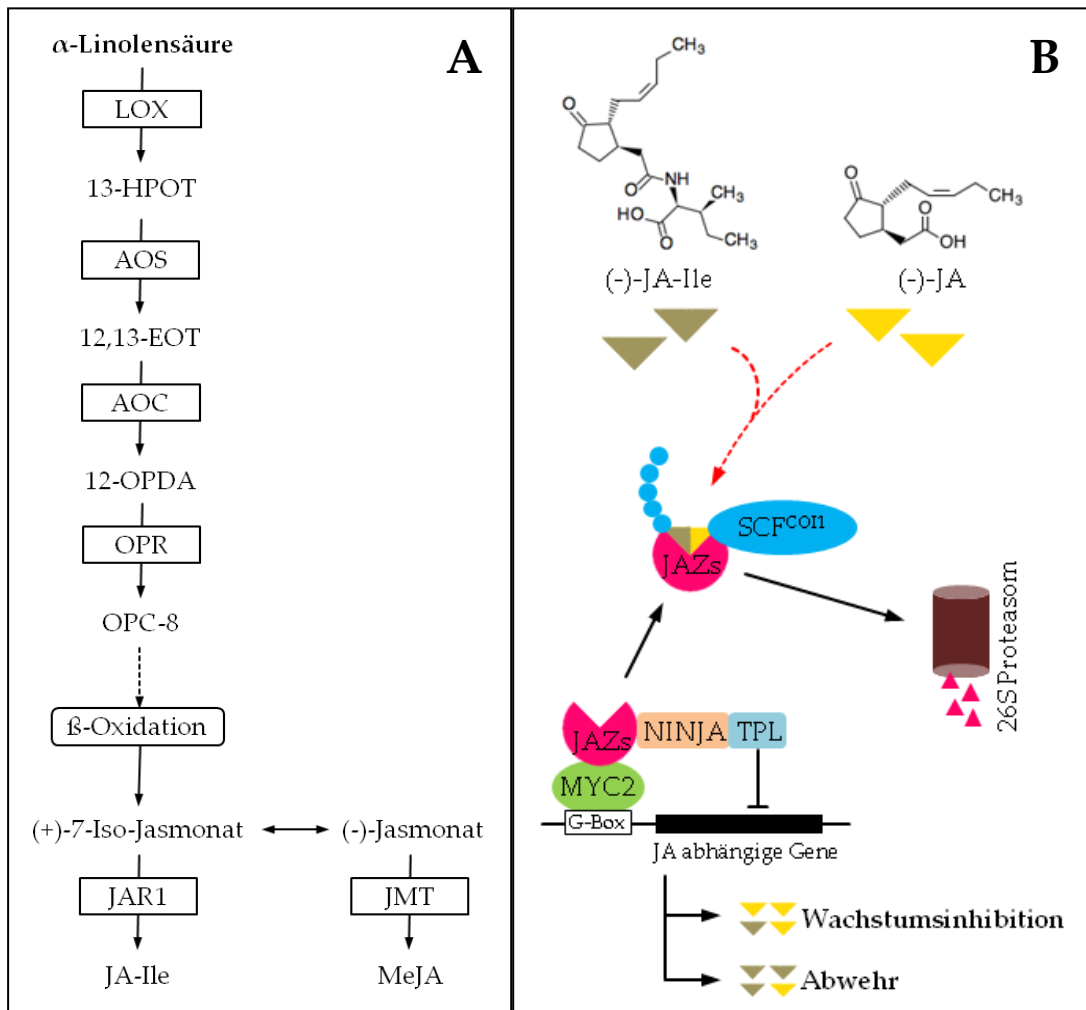


Abbildung 1.5: Biosynthese (A) und Signaltransduktion (B) von Jasmonaten (JAs). Detaillierte Beschreibungen der Abbildungen sind im Text (1.2.5) zu finden. **A** LOX – Lipoxygenase; AOS – Allene Oxide Synthase; AOC – Allene Oxide Cyclase; OPR – Oxophytodienoate Reductase; JAR1 – Jasmonate Resistant 1; JMT – Jasmonate O-Methyltransferase. Gestrichelte Pfeile stehen für mehrstufige Reaktionen. **B** JAZ – Jasmonate ZIM-Domain; NINJA – Novel Interactor of JAZ; TPL – Topless; MYC2 – Transcription factor MYC2.

1.2.6 ABSCISINSÄURE

Abcisinsäure (engl. Abscisic Acid, ABA) hat im Allgemeinen ähnlich wie JAs eine eher hemmende Wirkung auf das Pflanzenwachstum und spielt ebenfalls eine wichtige Rolle bei der Reaktion auf biotischen und abiotischen Stress [79]. Darüber hinaus ist ABA an der Aufrechterhaltung der Samen- und Knospenruhe und der Abscission von Blättern beteiligt, worauf auch die Namensgebung zurückgeht.

ABA ist ein ubiquitär synthetisiertes Sesquiterpenoid, das wie die Strigolaktone β -Carotin als Vorläufer hat (Abbildung 1.6 A). Dieses wird von der Familie der 9-cis-Epoxycarotenoid Dioxygenasen (NCED) zu Xanthoxin gespalten und durch ABA2 (Xanthoxin dehydrogenase) zu Abscisinaldehyd umgewandelt, welches von AAO3 (Abscisic Aldehyd Oxidase 3) zu ABA oxidiert wird [80]. Der Transport freier ABA oder inaktiver Konjugate (z.B. mit Glucose) findet im Xylem und Phloem statt.

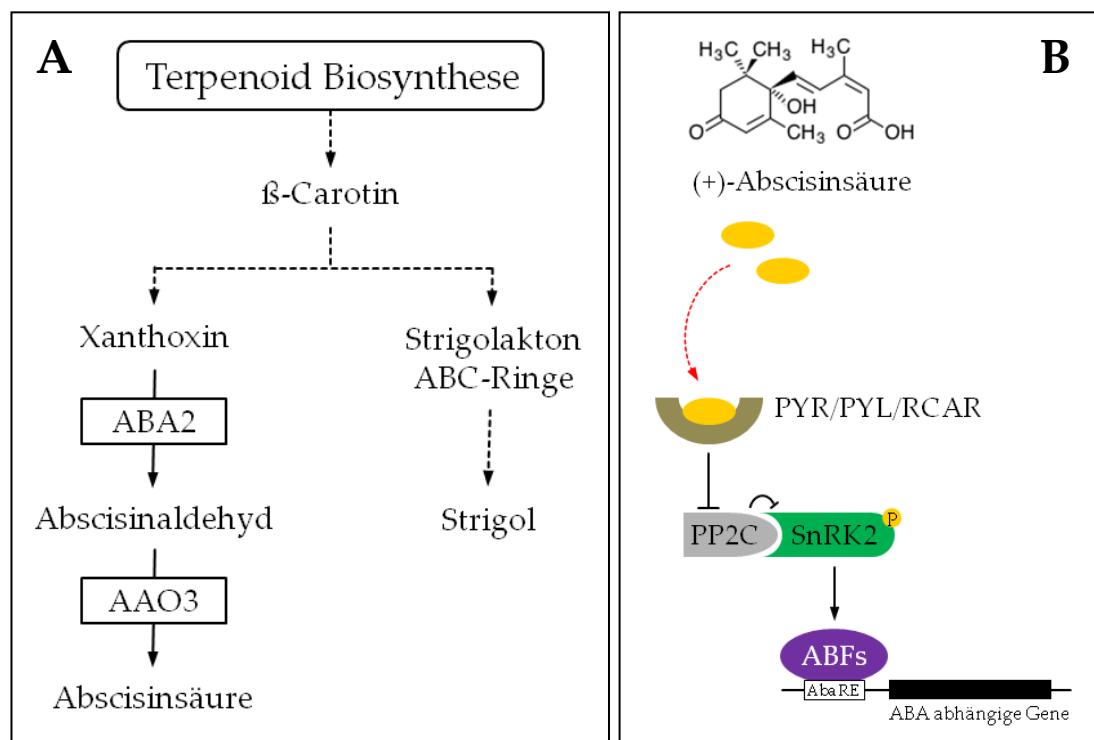


Abbildung 1.6: Biosynthese (A) und Signaltransduktion (B) von Abscisinsäure (ABA). Detaillierte Beschreibungen der Abbildungen sind dem Text (1.2.6) zu entnehmen. **A** ABA2 – Xanthoxin dehydrogenase; AAO3 – Abscisic Aldehyd Oxidase 3; Gestrichelte Pfeile stehen für mehrstufige Reaktionen. **B** PYR/PYL/RCAR – Pyrabactin Resistance/PYR-like/Regulatory Component of ABA Receptor; PP2C – Typ-2C-Proteinphosphatasen; SnRK2 – SnF1-Related Protein Kinase 2.

Die ABA-Signaltransduktion (Abbildung 1.6 **B**) findet über intrazelluläre Rezeptoren der PYR/PYL/RCAR-Familie (Pyrabactin Resistance/PYR-like/Regulatory Component of ABA Receptor) statt. Diese Interaktion führt zur Inaktivierung von Typ-2C-Proteinphosphatasen (PP2C), die in Abwesenheit von ABA SnRKs (SnF1-Related Protein Kinases) hemmen. Nach ABA-Bindung sind diese somit aktiv und können Transkriptionsfaktoren (ABA-responsive Element Binding Factors, ABFs) durch Phosphorylierung aktivieren, die wiederum die Expression ABA-abhängiger Gene steuern [81].

Da in Bezug auf Biosynthese, Signaltransduktion und Transport der hier vorgestellten Hormone und Wachstumsfaktoren auf Genebene bislang keine Erkenntnisse vorliegen, bietet sich für die umfassende Analyse der zahlreichen, damit assoziierten Gene eine Hochdurchsatzsequenzierung des gesamten Transkriptoms an (1.4). Zusätzlich schließt diese Art der Analyse auch Faktoren mit ein, deren Funktionen nicht offensichtlich mit phänotypischen Auswirkungen verknüpft oder gänzlich unbekannt sind.

1.3 KOLUMNARWACHSTUM BEIM APFEL

Das Kolumnarwachstum beim Apfel wurde erstmals in den frühen 1960er Jahren als eine natürliche Mutation der Sorte ‚McIntosh‘ beschrieben [82]. Dabei handelte es sich um eine somatoklonale Variation, die der aus Polen stammende Obstbauer Wijcik im Okanagan Valley (British Columbia, Kanada) entdeckte und die folglich nach ihm ‚McIntosh Wijcik‘ benannt wurde. Der kolumnare Phänotyp (Abbildung 1.7 **B**) ist eine extreme Form des Kurztriebwachses (spur type growth) [83, 84] und zeichnet sich durch verkürzte Internodien, einen größeren Stammdurchmesser und Fruchtspieße an Stelle von Seitentrieben aus [85]. Gelegentlich doch auftretende Seitentriebe wachsen in spitzem Winkel zum Hauptspross und dann annähernd parallel dazu [86]. Aufgrund dieser Eigenschaften können kolumnare Apfelbäume dichter als herkömmliche Apfelbäume (Abbildung 1.7 **A**) gepflanzt werden und steigern so den Ertrag pro Hektar um bis zu 300 % [87]. Weitere ökonomische Vorteile sind der minimale Aufwand für Rückschnitt und der Einsatz von Vollerntemaschinen [85]. Der Grad des kompakten Wuchses kann jedoch aufgrund von Umwelteinflüssen, dem Alter des Baumes [86, 88] sowie der Wahl der Unterlage [89] variieren.

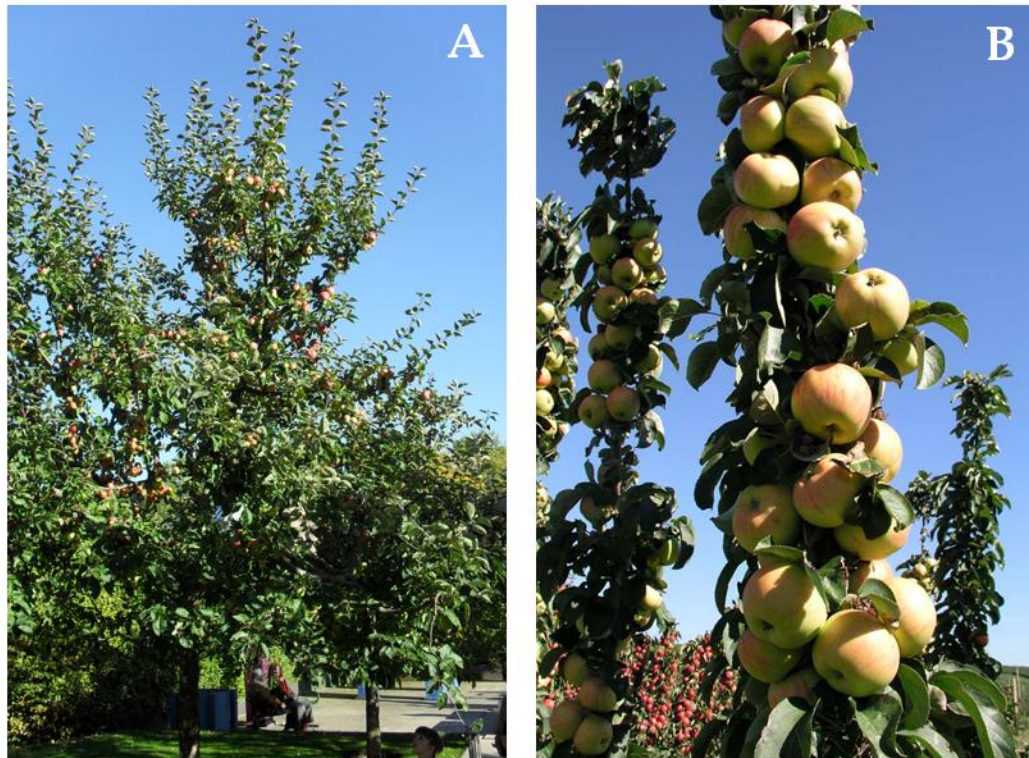


Abbildung 1.7: Vergleich der Wuchsformen. Gezeigt sind normalwüchsige (A) und kolumnare (B) Apfelbäume mit reifen Früchten. Die Bilder wurden am 10. Oktober 2010 im Palmengarten Frankfurt (A) und am 09. September 2004 in der Forschungsanstalt Geisenheim (B) aufgenommen.

Mehrere genetische Studien konnten durch Kreuzungen von kolumnaren und nicht kolumnaren Eltern zeigen, dass das Kolumnarwachstum von einem dominanten Allel des Gens *Columnar* (*Co*) kontrolliert wird [83, 90-92]. Da in Kreuzungen $co/co \times Co/co$ oft nicht das für Mendelsche Erbgänge typische 1:1 (kolumnar zu normalwüchsig) Verhältnis nachgewiesen werden kann, sind weitere genetische bzw. nichtgenetische Einflüsse nicht ausgeschlossen. Bislang schienen alle kolumnaren Sorten heterozygot für *Co* zu sein [93], jedoch ist es in der Forschungsanstalt (FA) Geisenheim kürzlich gelungen, eine für *Co* potentiell homozygote Sorte zu etablieren (Peter Braun, pers. Kommunikation).

Der Einsatz von molekularen Markern hat gezeigt, dass *Co* auf Chromosom 10 lokalisiert ist [93, 94]; eine präzise Lokalisation ist bislang nicht gelungen. Die entscheidende Chromosomenregion konnte mittels diverser Markeranalysen [93, 95-98] und mit Hilfe des sequenzierten Genoms [9] jedoch auf etwa 2 Mb eingegrenzt werden. Der Vergleich molekularer und kreuzungstechnischer Daten [99-102] ergab eine Relation von $1 \text{ cM} = 0,399 \text{ Mb}$. Danach ergibt sich eine Region

von 17 bis 19 Mb auf Chromosom 10, in der der potentielle *Co*-Locus positioniert sein muss (Abbildung 1.8, oben). Weiterhin scheint zumindest ein Teil von Chromosom 3 ebenfalls mit dem *Co*-Gen gekoppelt zu sein (Abbildung 1.8, unten). Die genaue Kartierung und molekulare Analyse dieser Region ist derzeit Gegenstand einer weiteren Dissertation [103]. Die Isolierung des *Co*-Gens ist von wirtschaftlichem Interesse, um dieses eventuell gentechnisch in bereits marktgängige Apfelsorten zu überführen. Damit könnten kolumnare Eigenschaften (Wuchsform) gezielt in erfolgreiche Sorten transferiert werden, ohne sonstige Eigenschaften wie Fruchtqualität und Schädlingsresistenz zu verändern.

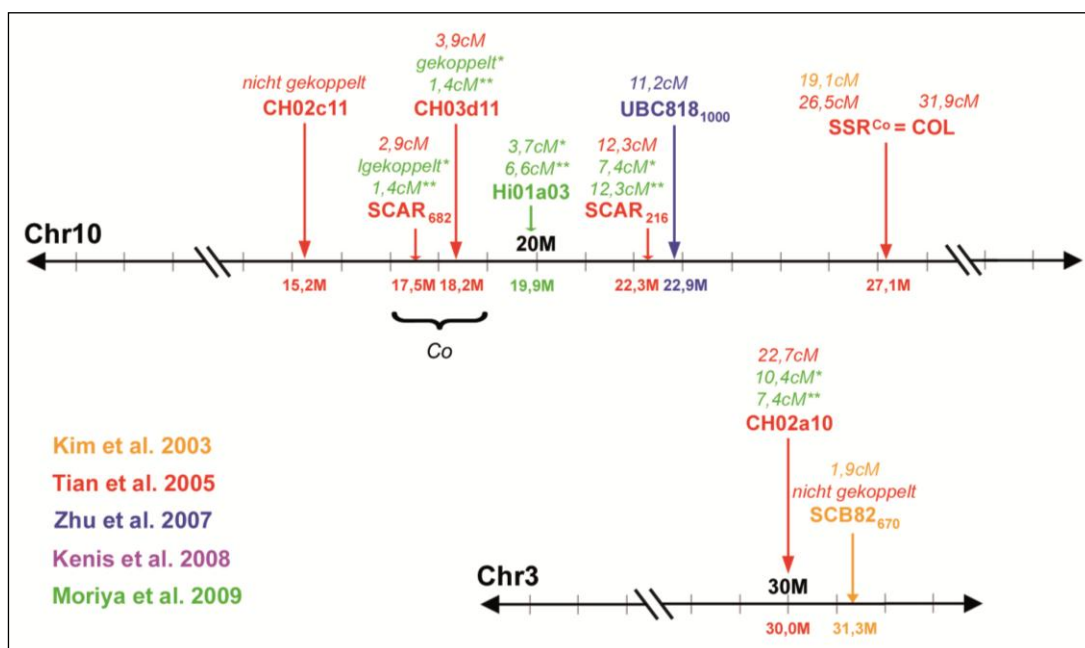


Abbildung 1.8: Lokalisation des *Co*-Gens. Gezeigt sind die zur Bestimmung der Lage des *Co*-Locus verwendeten Marker (die farbliche Markierung entspricht der jeweiligen Publikation) und ihre chromosomale Lokalisation. Daraus ergibt sich eine 2 Mb große Region von 17 bis 19 Mb auf Chromosom 10. Die Namen der Marker sind fett, die genetischen Distanzen (Rekombinationsfrequenzen) kursiv dargestellt. */** markieren Werte, die an verschiedenen Sorten gemessen wurden. M = Mb – Megabasen; Chr – Chromosom.

Das Wissen um die molekularen Hintergründe des Kolumnarwachstums ist dagegen noch recht rudimentär. So sind weder Sequenz, Genprodukt, Funktion noch mögliche Interaktionen bekannt. Dagegen gibt es auf dem Gebiet des Phytohormonhaushalts kolumnarer Apfelbäume einige gut untersuchte Erkenntnisse, weil deren Beteiligung an der Entstehung dieses ungewöhnlichen

Phänotyps offensichtlich erscheint. So konnte mehrfach gezeigt werden, dass die Konzentration von freien Auxinen und Cytokinin in Sprossspitzen kolumnarer (verglichen mit normalwüchsigen) Bäumen wesentlich höher ist [92, 104-106]. In Sorten mit erhöhtem Kurztriebwuchs (s.o.) bei Abwesenheit des *Co*-Gens wurden dagegen niedrigere Auxinkonzentrationen gemessen [107, 108]. Diese Unterschiede in den Hormonkonzentrationen führen wahrscheinlich dazu, dass kolumnare Bäume („Maypole“ und „Tuscan“) über die gesamte Wuchsperiode bis Oktober wachsen, während „McIntosh“ sein Wachstum im Juli einstellt [105]. Darüber hinaus konnte gezeigt werden, dass kolumnare Bäume neben den erhöhten, endogenen Cytokininspiegeln eine ausgeprägte Toleranz gegenüber exogenem Cytokinin aufweisen [109, 110]. Die endogenen Konzentrationen von Abscisinsäure und polaren Gibberellinen sind dagegen reduziert [92, 111]. Abbildung 1.9 fasst diese bisherigen Erkenntnisse zusammen.

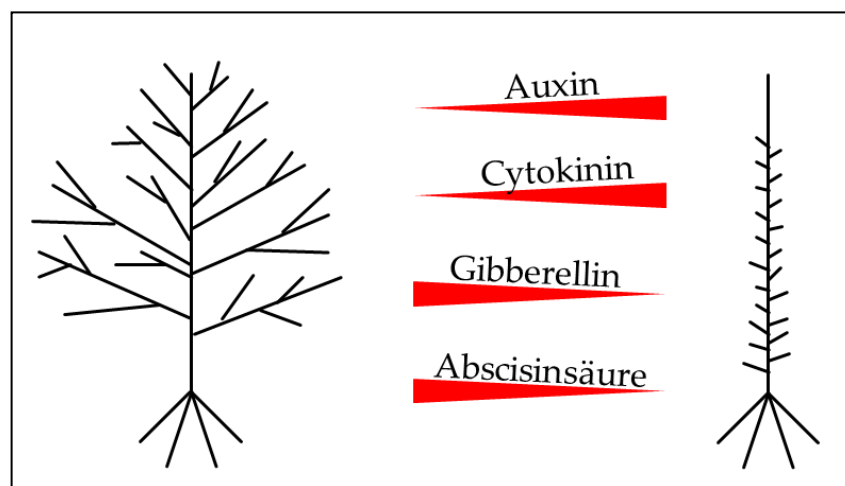


Abbildung 1.9: Hormonale Situation in kolumnaren Bäumen. Auxin und Cytokinin sind in Sprossspitzen kolumnarer Bäume (rechts) höher, Gibberellin und Abscisinsäure niedriger konzentriert als in normalwüchsigen Bäumen (links).

Eine aktuelle Studie aus dem Jahr 2009 [17] beschreibt einen kolumnaren Phänotyp hervorgerufen durch die Überexpression des floralen Identitätsgens *LEAFY* (*LFY*) [17, 112], allerdings weist dieser Phänotyp nicht alle der typischen Charakteristika kolumnarer Bäume auf (Peter Braun, persönliche Kommunikation).

1.4 GENEXPRESSIONSANALYSE MIT HILFE VON NEXT GENERATION SEQUENCING (NGS)

Mit dem Aufkommen und der kontinuierlichen Verbesserung von Sequenzier-Technologien der zweiten (next) Generation seit Mitte des vergangenen Jahrzehnts [113] steht mit RNA-Seq [114] heute eine revolutionäre Methode für die umfassende Analyse ganzer Transkriptome zur Verfügung. Wie auch bei *de novo*- oder Resequenzierungen ganzer Genome liegt der Vorteil der Verwendung von NGS hierbei im extrem hohen Durchsatz.

Derzeit existieren drei konkurrierende Plattformen, die sich am Markt trotz einer großen Anzahl neuartiger Entwicklungen (Third Generation Sequencing) durchgesetzt haben: Die 454/FLX-Technologie von Roche als erste eingeführte der neuen Plattformen beruht auf der Detektion von Pyrophosphat, das beim Einbau von Nukleotiden abgespalten wird [115]. Um ein ausreichend starkes Signal zu erhalten, wird jedes DNA-Molekül an die Oberfläche einer winzigen Kugel (Bead) gebunden und per Emulsions-PCR (emPCR) in einem Mikroreaktionsraum amplifiziert [116]. Die eigentliche Sequenzierung erfolgt dann in einer Picotiterplatte durch sukzessives, wiederholtes Überspülen mit den vier Nukleotiden. So können etwa eine Millionen Sequenzen mit einer Länge von bis zu 700 nt (entspricht 700 Mb) in einem 23-Stunden-Lauf generiert werden.

Die SOLid™-Technologie von Applied Biosystems/Life Technologies bedient sich ebenfalls einer emPCR; die Beads werden anschließend jedoch auf einem Glaträger immobilisiert. Die Sequenzierung erfolgt durch Anlagerung degenerierter und fluoreszenzmarkierter Oligonukleotide (Oktamere), die aufgrund der ersten beiden Basen spezifisch am zu sequenzierenden Molekül hybridisieren. Die letzten drei Nukleotide werden mit der Fluoreszenzgruppe nach Ligation abgespalten, bevor das nächste Oktamer hybridisieren kann. Da über diese Art der Kodierung noch nicht die Sequenz bestimmt werden kann (ein Fluoreszenzfarbstoff wird für jeweils vier verschiedene Dinukleotidkombinationen benutzt), wird dieser Prozess insgesamt fünfmal von Position n bis $n-5$ wiederholt, um jede Basenposition zweimal zu sequenzieren, was in einer sehr geringen Fehlerrate resultiert. Bei einem 7-tägigen Lauf können mit diesem System so bis zu 300 Gb sequenziert werden, wobei aber nur Sequenzlängen von 50-75 nt erreicht werden, die im Anschluss einen hohen bioinformatischen Aufwand erfordern.

Den höchsten Durchsatz erzielen mit bis zu 600 Gb und 150 nt Sequenzlänge jedoch die beiden Illumina-Systeme Genome Analyzer IIx und HiSeq2000. Hierbei werden die DNA-Moleküle in Vorbereitung auf die Sequenzierung an zwei verschiedene, oberflächengekoppelte Oligonukleotide auf einer Flow Cell hybridisiert und per Bridge Amplification zu sogenannten Clustern vervielfältigt [117]. Dazu sind spezifische Adapter an den Enden der Moleküle notwendig, die es darüber hinaus ermöglichen, das Molekül von beiden Seiten zu sequenzieren. Um dies zu realisieren, werden alle Fragmente eines Clusters durch Restriktion in einem der beiden oberflächengekoppelten Oligonukleotide einheitlich ausgerichtet. Anschließend erfolgt die Neusynthese des Gegenstranges, wobei nur Nukleotide mit reversiblen, fluoreszenzmarkierten Terminatoren eingebaut werden [118]. Bei deren Abspaltung erfolgt die Bestimmung der eingebauten Base anhand eines Vier-Farben-Systems. Nach erfolgreicher Sequenzierung können die Cluster regeneriert werden, das jeweils andere Oberflächenoligonukleotid kann restringiert und die Moleküle können von der anderen Seite her sequenziert werden.

In Anbetracht dieser leicht herzustellenden, gigantischen Datenmengen rücken klassische und ebenfalls auf hohen Durchsatz ausgerichtete Verfahren wie die Generierung von cDNA-Banken (Expressed Sequence Tag (EST)-Sequenzierung [119]), SAGE (Serial Analysis of Gene Expression [120]) oder Microarrays [121, 122] zusehends in den Hintergrund [123, 124]. Die beiden erstgenannten Methoden beruhen auf der Sequenzierung (Identifizierung) und Quantifizierung von je einer langen bzw. vieler kurzer, aus RNA umgeschriebenen cDNAs. ESTs klassischer cDNA-Banken sind dabei aufgrund der längeren Sequenzen (eine Sequenz = ein EST) leichter zu identifizieren, während bei der SAGE-Technologie zwar ein höherer Durchsatz erreicht wird (eine Sequenz = viele ESTs), die einzelnen cDNA Abschnitte (Tags) mit Längen von 10-14 nt aber schwerer zu identifizieren sind. Microarrays beruhen dagegen nicht auf Sequenzierung, sondern auf Sonden-Hybridisierung. Dies hat den Vorteil, die Regulation tausender Gene in einem Experiment zu bestimmen, allerdings müssen dazu die Zielsequenzen bekannt sein. Spleiß-Varianten und unbekannte Gene sind so nur schwer nachweisbar, darüber hinaus ist lediglich eine relative Quantifizierung möglich. Während Methoden wie SAGE unter Einbeziehung von NGS-Technologien kontinuierlich weiterverbessert werden (SuperSAGE [125, 126]), werden Microarrays gerne zur Bestätigung von RNA-Seq Ergebnissen eingesetzt.

Das Anwendungsspektrum von RNA-Seq ist dagegen sehr breit gefächert, bedarf nur geringen Laboraufwands und setzt keinerlei Kenntnisse des vorliegenden Transkriptoms voraus, weshalb die Technologie in Modell- wie auch in Nicht-Modellorganismen Anwendung findet. So lassen sich Genstrukturen, alternative Spleiß-Ereignisse und Expressionsstärke von Transkripten in der Entwicklung von Maisblättern genauso analysieren wie die Reaktion auf osmotischen Stress und ABA in *Sorghum bicolor* [127, 128]. Beschrieben ist weiterhin die erfolgreiche Suche nach Kandidatengen für verschiedene Stoffwechselwege in wirtschaftlich interessanten Pflanzen. Dabei lassen sich putative, in der Ginsenosid-Biosynthese involvierte Gene im Ginseng (*Panax quinquefolius*) genauso nachweisen wie Stoffwechselwege von Tee (*Camellia sinensis*)-spezifischen Komponenten [129, 130]. Ebenfalls erfolgreich beschrieben sind Vergleiche verschiedener Genotypen in Bezug auf Infektionen, Blüten- und Fruchtentwicklung bei Kastanie (*Castanea*), Buchweizen (*Fagopyrum*) und Olive (*Olea*) [131-133].

1.5 ZIELSETZUNG

Das Ziel der vorliegenden Arbeit ist die komparative Analyse der Transkriptome von kolumnaren und normalwüchsigen Apfelbäumen (*Malus x domestica*) anhand der in der Forschungsanstalt Geisenheim gezüchteten Sorten A14-190-19 (normalwüchsig) und Procats28 (kolumnar). Dabei sollen Unterschiede in der Genexpression qualitativ und quantitativ charakterisiert und wenn möglich in Bezug zum kolumnaren Phänotyp gesetzt werden. Die naheliegende Hypothese ist hierbei, dass durch die Wirkung des dominanten *Co*-Allels die Expression etlicher Gene deutlich beeinflusst werden muss, um die massiven Änderungen im Habitus kolumnarer Bäume zu erzeugen. Sollte sich diese Hypothese bewahrheiten, ist ein weiteres Ziel, aus Regulation, Funktion oder Lokalisation der entsprechenden Gene Rückschlüsse auf die Funktion des Genprodukts zu ziehen. Bislang gibt es dazu noch keine konkreten Erkenntnisse.

Da das oberirdische Wachstum von Pflanzen maßgeblich von Faktoren bestimmt wird, die im Sprossapikalmeristem (SAM) gebildet werden, soll dieses Gewebe als Untersuchungsobjekt verwendet werden. Aufgrund der Tatsache, dass zu verschiedenen Zeitpunkten der vegetativen Wachstumsperiode unterschiedliche Gene aktiv sind, sollen Proben diverser Entnahmezeitpunkte betrachtet werden.

Dies soll vor allem dazu beitragen, diese temporär aktivierten/reprimierten Gene von solchen separieren zu können, die aufgrund der Wirkung des *Co*-Allels in ihrer Expression verändert werden. Mittels NGS-Analyse sollen daher die Transkriptome kolumnarer und nicht-kolumnarer SAM umfassend untersucht werden. Die Anwendung der NGS-Technologie im Hinblick auf solche Fragestellungen ist relativ neu und befindet sich noch in der Entwicklung. Daher soll in dieser Arbeit auch die Methodik der Datenerzeugung und der bioinformatischen Auswertung etabliert und verbessert werden.

Die auf diese Weise gewonnen und verarbeiteten Daten sollen zielgerichtet im Hinblick auf Wachstum und Wachstumssteuerung analysiert werden, wobei ein Fokus auf Gene gelegt wird, welche am Phytohormonhaushalt der Pflanze beteiligt sind. Zum Teil sollen die Ergebnisse durch weitere Expressionsstudien mit alternativen Methoden (Microarrays und qRT-PCR) überprüft werden. Die Arbeit ist ein Teilprojekt des vom Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz geförderten Projekts (Nr. 511-06.01-28-1-43.042-07) zur Identifikation und Funktionsaufklärung des *Co*-Gens.

2 MATERIAL UND METHODEN

2.1 PFLANZENMATERIAL

Für alle in dieser Arbeit aufgeführten Untersuchungen wurden ausschließlich die nicht-kolumnare Apfelsorte (*Malus x domestica*) A14-190-93 und die kolumnare Apfelsorte Procats28 (im Folgenden als A14 und P28 bezeichnet) verwendet. A14 ist eine Kreuzung der Sorten GOLDEN DELICIOUS ‚Weinsberg‘ (nicht-kolumnar) und WALTZ (kolumnar), wobei auf nicht-kolumnare Nachkommen selektiert wurde. P28 wurde aus der kolumnaren Sorte FLAMENCO und der nicht-kolumnaren Sorte TOPAZ gekreuzt. Die Entnahme von Triebspitzen erfolgte zu verschiedenen Zeitpunkten (Tabelle 2.1) an Bäumen, die in der Forschungsanstalt (FA) Geisenheim gezüchtet wurden (Freilandmaterial). Dazu wurden die Triebe etwa 2 cm unterhalb des Sprossapikalmeristems (shoot apical meristem, SAM) dekapitiert. Die so gewonnenen Triebspitzen wurden auf Eis lichtdicht nach Mainz transportiert und bei -80°C bis zur Gewinnung der RNA (siehe 2.2.1) tiefgefroren.

Tabelle 2.1: Ausgangsmaterial. Übersicht über die Entnahmezeitpunkte und spätere Verwendung von Triebspitzen nicht-kolumnarer A14 und kolumnarer P28 aus der FA Geisenheim. NGS – Next Generation Sequencing; MA – Microarray.

Zeitpunkt der Entnahme	A14	P28	Verwendung
22. Mai 2009	✓	✓	NGS, MA
29. September 2009		✓	NGS
6. April 2010	✓	✓	MA
20. April 2010	✓	✓	MA
4. Mai 2010	✓	✓	MA
20. Juli 2010	✓	✓	NGS

Weiterhin wurden ebenfalls in der FA Geisenheim etablierte *in vitro*-Kulturen von A14 und P28 zur Gewinnung von RNA genutzt. Diese werden ganzjährig aus Sprossspitzenexplantaten regeneriert und auf MS Medium [134] kultiviert. Dem Medium werden weiterhin Sorbit (30 g/L), Glucuronsäure (1,3 g/L), Benzylaminopurin (BA, 1 mg/L) und Indol-3-Buttersäure (IBA, 0,1 mg/L) zugegeben. Die Kultivierung erfolgt bei 22°C und 14 h Photoperiode.

2.2 MOLEKULARBIOLOGISCHE METHODEN

2.2.1 RNA-EXTRAKTION

Die Triebspitzen des Freilandmaterials wurden zunächst grob von Spross- und Blattgewebe (Lamina und Petiolus) befreit und das SAM wurde so weit wie möglich freipräpariert. Das gewonnene Material wurde mit Hilfe von flüssigem Stickstoff in einem Mörser homogenisiert und unmittelbar nach Abdampfen des Stickstoffs in den Lysepuffer (siehe unten) überführt. Die Präparation der RNA erfolgte mittels des innuPREP Plant RNA Kits (Analytik Jena, Jena, Deutschland) nach Angaben des Herstellers. Verwendet wurde ausschließlich Lysepuffer PL; die anschließende Inkubation erfolgte in einem Thermomixer (Eppendorf, Hamburg, Deutschland) bei 300 rpm und RT. Die *in vitro*-Kulturen wurden vor der RNA-Extraktion von Kallusgewebe befreit und ohne weitere Präparation der oben beschriebenen Aufarbeitung zugeführt.

2.2.2 DNASE-VERDAU

Die in 2.2.1 gewonnene RNA wurde zur Entfernung von DNA-Rückständen mit DNase I verdaut. Dazu wurde die RNA-Lösung mit 0,1 Vol. 10 x Puffer und 1 µl DNase I (2 U/µl, beides NEB, Ipswich, USA) versetzt und bei 37°C für 30 min inkubiert. Inaktivierung des Enzyms erfolgte unter Zusatz von 0,1 Vol. 50 mM EDTA für 15 min bei 65°C.

2.2.3 ETHANOLFÄLLUNG

Zur Entfernung von Salzen und zur Aufkonzentrierung der Nukleinsäure wurde die DNase-verdaute RNA in saurem Milieu präzipitiert. Dazu wurde die Lösung mit RNase-freiem Wasser auf mindestens 100 µl aufgefüllt, mit 0,1 Vol. 3 M Natriumacetat (pH 5,2) und 2,5 Vol. eiskaltem Ethanol (EtOH abs.) versetzt, invertiert und für mehr als 45 min bei -20°C inkubiert. Zum Pelletieren der RNA wurde die Lösung für 45 min bei 4°C und 14000 g zentrifugiert, der Überstand abgezogen und das Pellet mit 500 µl 70 % Ethanol gewaschen. Nach erneuter Zentrifugation für 10 min bei 14000 x g (RT) wurde der Überstand wieder vollständig abgenommen und das Pellet in einer Vakuumzentrifuge getrocknet. Die RNA wurde anschließend in einem geeigneten Volumen RNase-freien Wassers (meist 20 µl) gelöst. Die cDNAs wurden ebenfalls präzipitiert, hier wurde jedoch 3 M Natriumacetat mit pH 7,2 eingesetzt, um ein neutrales Milieu für die Fällung zu gewährleisten. Alle weiteren Schritte wurden wie oben beschrieben durchgeführt.

2.2.4 POLY-A⁺-AUFREINIGUNG

Die Anreicherung der mRNA aus der extrahierten Total-RNA wurde mit Hilfe des NucleoTrap® mRNA Mini Kits (Macherey Nagel, Düren, Deutschland) durchgeführt. Dabei wurde ein modifiziertes Protokoll ohne den Einsatz der NucleoTrap® Mikrofilter angewandt. Die gefällte RNA (aus 2.2.3) wurde entweder direkt in 500 µl Puffer RM1 gelöst oder die schon in RNase-freiem Wasser aufgenommene RNA wurde auf 250 µl aufgefüllt und mit 1 Vol. Puffer RM0 versetzt. Die Lösung wurde mit 15 µl oligo(dT) Latex Beads versetzt, bei 68°C 5 min und bei RT 10 min inkubiert und danach 15 s bei 2000 x g an- und 5 min bei 11000 x g abzentrifugiert. Der Überstand wurde abgenommen und das Pellet insgesamt dreimal gewaschen (1 x 600 µl Puffer RM2, 2 x 500 µl Puffer RM3). Nach jedem Waschschrift wurde wie oben beschrieben zentrifugiert und der Überstand abgezogen. Im letzten Schritt wurde das getrocknete Pellet in 30 µl RNase-freiem Wasser aufgenommen und bei 68°C 7 min inkubiert. Nach Zentrifugation (4 min, 11000 x g, RT) wurde der Überstand mit der gelösten RNA überführt und das Pellet verworfen. Die Qualität der poly-A⁺-Aufreinigung wurde mit Hilfe des Agilent 2100 Bioanalyzers (siehe 2.2.7) bestimmt und gegebenenfalls (bei zu hohem Anteil ribosomaler RNAs) eine zweite poly-A⁺-Aufreinigung durchgeführt.

2.2.5 RNA-AMPLIFIKATION

Aufgrund unzureichender Mengen an poly-A⁺-aufgereinigter mRNA für die Folgeanwendungen wurde eine Amplifikation mit Hilfe des MessageAmp[™] II aRNA Kits (Ambion, Austin, USA) durchgeführt. Verfahren wurde dabei nach Angaben des Herstellers, wobei die maximale Menge mRNA eingesetzt (11 µl) und die *in vitro*-Transkription insgesamt für 14 h inkubiert wurde. Die auf diese Weise gewonnene aRNA (antisense RNA) wurde in 100 µl RNase-freiem Wasser eluiert und Ausbeute und Qualität wurden auf dem Bioanalyzer (2.2.7) bestimmt.

2.2.6 AGAROSE-GELELEKTROPHORESE

Für die Größenauftrennung der verschiedenen RNAs (Total-RNA, mRNA, aRNA) wurden denaturierende Agarosegele (1,2 % Agarose in 1 x MOPS, versetzt mit 15 % Formaldehyd) verwendet. Als Laufpuffer diente ebenfalls 1 x MOPS. Die Proben wurden wie folgt vorbereitet: 1 Vol. RNA-Lösung (mindestens 3 µl), 1 µl Ethidiumbromid-Stammlösung und 3 Vol. RNA-Ladepuffer wurden bei 72°C für 10 min denaturiert und mit 0,1 Vol. RNA-Färbelösung versetzt. Die Elektrophorese des Gels (7 cm x 6 cm x 0,1 cm) erfolgte in GENTerphorese[™]-Gelkammersystemen (GENTerprise, Mainz, Deutschland) bei 130 mA für 20-30 min. Im Anschluss wurden die Gele auf einem UV-Transilluminator bei einer Wellenlänge von 312 nm betrachtet und mit Hilfe des Carestream GelLogic 112 Systems (Carestream, Health, Rochester, USA) dokumentiert.

Die Auftrennung von cDNAs erfolgte dagegen auf nicht-denaturierenden Agarosegelen (0,8 % - 3 % Agarose in 1 x TBE), als Laufpuffer diente ebenfalls 1 x TBE. Auch hierbei wurde das oben genannte GENTerphorese[™]-Gelkammersystem sowie das horizontale Gelelektrophoresesystem Gibco BRL Sunrise[™] 96 (Life Technologies, Carlsbad, USA) eingesetzt. Die Proben wurden mit 1/6 Vol. DNA-Ladepuffer versetzt. Nach der Elektrophorese wurden die Gele in einer Ethidiumbromid-Färbelösung (0,0005 % Ethidiumbromid in 1 x E-Puffer) gefärbt und wie oben beschrieben dokumentiert. Als Molekulargewichtsmarker diente GeneRuler[™] 100bp Plus DNA Ladder (Fermentas, St. Leon-Roth, Deutschland).

2.2.7 QUALITÄTS- UND KONZENTRATIONSBESTIMMUNG DER RNA

Zur Bestimmung von Reinheit, etwaiger Degradation und Konzentration wurden alle gewonnenen RNAs mit Hilfe des Agilent 2100 Bioanalyzers (Agilent, Santa Clara, USA) analysiert. Die Integrität von Total-RNA wurde dabei mit dem RNA 6000 Nano Kit bestimmt, während mRNA und aRNA anhand des RNA 6000 Pico Kits (beides Agilent) untersucht wurden. Alle cDNAs wurden ebenfalls einer eingehenden Qualitätskontrolle unterzogen; hierfür wurden Chips aus dem High Sensitivity DNA Kit (Agilent) verwendet. Alle Untersuchungen wurden nach Herstellerprotokollen durchgeführt.

Zur Bestätigung der gemessenen Konzentrationen wurden weiterhin der NanoDrop ND-1000 (PeqLab, Erlangen, Deutschland) und das Qubit Fluorometer (Invitrogen, Carlsbad, USA) eingesetzt.

2.2.8 cDNA-SYNTHESE

Sowohl für die Sequenzierungen als auch für die qRT-PCR wurden die verschiedenen RNAs in cDNA umgeschrieben. Für alle Synthesen wurde ausschließlich SuperScript III Reverse Transkriptase (Invitrogen, Carlsbad, USA) nach Angaben des Herstellers eingesetzt.

2.2.8.1 Sequenzierungen

Für alle Sequenzierungen (vgl. 2.3) wurde ausschließlich doppelsträngige cDNA (ds cDNA) als Ausgangsmaterial für die Erstellung der Bibliotheken vorausgesetzt. Ausgehend von amplifizierter aRNA wurden daher für die Erststrangsynthese (2.2.8) 5 μ M Random Hexamer Primer (Fermentas, St. Leon-Roth, Deutschland) eingesetzt. Zweitstrangsynthese und Endreparatur wurden nach dem Fermentas Zweitstrang cDNA Synthese Protokoll durchgeführt (<http://fermentas.com/en/support/Application-Protocols/>), wobei die dafür notwendigen Enzyme RNase H (5 U/ μ l), DNA Polymerase I (10 U/ μ l) und T4 DNA Polymerase (5 U/ μ l) ebenfalls von Fermentas bezogen wurden. Die auf diese Weise hergestellten cDNAs wurden anschließend vereinigt, gefällt (2.2.3) und auf dem Agilent 2100 Bioanalyzer (2.2.7) gemessen.

2.2.8.2 qRT-PCR

Da für alle qRT-PCR-Experimente Total-RNA aus *in vitro*-Kulturen die Grundlage darstellte, wurde für die Erststrangsynthese (2.2.8) 50 μM oligo(dT)₂₀ Primer eingesetzt. Jede Synthese (A14 und P28) basierte auf 800 ng RNA, wobei als exogener Standard jeweils 100 ng RNA aus der humanen Brustkarzinom-Zelllinie MDA-MB468 hinzugegeben wurde. Um die Vergleichbarkeit der Proben zu gewährleisten, wurden nur RNAs mit identischer Integrität (RIN) $\geq 8,0$ [135, 136] und Mastermixe zur Verhinderung von Pipettierungenauigkeiten eingesetzt. Die Qualität der so erhaltenen cDNAs wurde anschließend per PCR (2.2.9) mit GAPDH-spezifischen Primern (für *Malus x domestica* und *Homo sapiens*) getestet. Die positiv getesteten cDNAs wurden daraufhin 1:4 mit HPLC-Wasser verdünnt und bei -20°C gelagert.

2.2.9 PCR UND AUFREINIGUNG

Alle Primer (Invitrogen, Carlsbad, USA) wurden mit HPLC-H₂O auf 10 pmol/ μl (10 μM) eingestellt und per PCR getestet. Die Reaktionen wurden in 50 μl -Ansätzen mit 1 μl der verdünnten cDNA (2.2.8.2), je 10 pmol Primer, 0,2 mM dNTPs und 0,3 U GoTaq® DNA Polymerase (Promega, Madison, USA) in einem Tprofessional Thermocycler (Biometra, Göttingen, Deutschland) durchgeführt: Initiale Denaturierung 5 min 95°C , gefolgt von 55 Zyklen mit 30 sec 95°C , 30 sec 60°C , 30 sec 72°C und finaler Elongation für 10 min bei 72°C . Erfolgreich amplifizierte Fragmente wurden für Folgeanwendungen (Sequenzierung) mit 10 U Exonuklease I und 0,9 U Shrimp Alkalischer Phosphatase (Fermentas, St. Leon-Roth, Deutschland) verdaut.

2.2.10 KLONIERUNG VON PCR-FRAGMENTEN

100 ng der aufgereinigten PCR-Fragmente (2.2.9) wurden in 6 μl -Ansätzen in 5 ng pGEM®-T Easy Vektor (Promega, Madison, USA) ligiert (1 U T4 DNA Ligase (Roche, Indianapolis, USA), 4°C über Nacht). 1-2 μl Ligationsansatz wurden in elektrokompente Zellen des Bakterienstammes DH10B [137] transformiert [138], bei 37°C für 45 min inkubiert und auf Amp/X-Gal/IPTG-haltige Agarplatten ausgestrichen (37°C , ü. N.). Potentiell rekombinante Kolonien wurden gepickt und in LB-Medium [139] mit 1 % Ampicillin-Stammlösung bei 37°C ü. N. inkubiert.

2.2.11 PLASMID-PRÄPARATION

Die Isolation der rekombinanten Plasmide aus den über-Nacht-Kulturen (2.2.10) erfolgte mit Hilfe des peqGOLD Plasmid Miniprep Kits I (PeqLab, Erlangen, Deutschland) nach Herstellerangaben. Zur Bestimmung der Integratgröße wurden die Plasmide mit EcoRI (NEB, Ipswich, USA) restringiert und gelelektrophoretisch aufgetrennt (2.2.6).

2.2.12 SANGER-SEQUENZIERUNG

Zur Bestätigung von PCR-Produkten und erfolgreich klonierten Fragmenten (richtige Fragmentlänge) wurden diese sequenziert. Dazu wurden 50-100 ng der aufgereinigten PCR-Produkte mit genspezifischem Primer und 400-700 ng der aufgereinigten Plasmide mit T7- oder SP6-Primer bei der Firma StarSEQ GmbH (Mainz, Deutschland) sequenzieren lassen.

2.2.13 MICROARRAY-ANALYSEN

Alle Microarray-Analysen wurden vollständig von der Firma febit (Heidelberg, Deutschland) durchgeführt. Für 1500 ausgewählte Gene [SI Fasta_5] wurden je zwei 50 nt große Sonden synthetisiert. Als Ausgangsmaterial diente Total-RNA (jeweils A14 und P28) von insgesamt vier Zeitpunkten (Tabelle 2.1), wobei nur die differentielle Genexpression von A14 und P28 zueinander analysiert wurde.

2.2.14 QUANTITATIVE REAL-TIME-PCR

Alle qRT-PCRs erfolgten in 10 µl-Ansätzen in 96 Well PCR Platten (Biozym, Hessisch Oldendorf, Deutschland). Als Master Mix diente Power SYBR® Green PCR Master Mix (Applied Biosystems, Carlsbad, USA) mit je 0,3 µM (3 pmol) der beiden genspezifischen Primer. Zu je 9 µl Master Mix wurden die verdünnten cDNAs (2.2.8.2) in Triplets à 1 µl pipettiert, wobei für jedes getestete Gen eine Negativkontrolle (HPLC-Wasser) mitgeführt wurde. Die Platten wurden mit BZO Adhesive Optical Seal Film (Biozym) versiegelt und kurz abzentrifugiert. Alle Läufe

erfolgten in einem Applied Biosystems 7500 Fast Real-Time PCR System mit dem folgenden Programm: Initiale Denaturierung 95°C für 10 min, gefolgt von 45 Zyklen mit 15 sec 95°C, 30 sec 60°C und 30 sec 72°C, abschließend zwei Zyklen zur Messung der Dissoziation mit 15 sec 95°C, 15 sec 60°C. Die Datenaufnahme erfolgte in der Elongationsphase (72°C). Als exogener Standard wurde auf jeder Platte humanes GAPDH mitgeführt.

2.2.15 KEIMUNG VON APFELKERNEN

Da anfangs kein keimungsfähiges Material von A14 oder P28 zur Verfügung stand, wurde für die Chromosomenpräparate auf handelsübliche Kulturäpfel zurückgegriffen. Die Kerne aus reifen Äpfeln wurden ca. eine Woche trocken gelassen, zum Schutz vor Schimmel mit 70 % Ethanol desinfiziert und in Petrischalen auf feuchtem Handtuchpapier lichtdicht bei 4°C mindestens vier Wochen stratifiziert. Bei einsetzender Keimung wurden die Petrischalen auf RT gebracht und der Status der Keimung wurde bis zur Entnahme (2.2.16) täglich kontrolliert.

2.2.16 PROTOPLASTENISOLIERUNG

Als Ausgangsmaterial für die Protoplastenisolation dienten 1-2 mm lange Wurzelspitzen, die den 2-3 Tage alten Keimlingen (2.2.15) entfernt wurden. Diese wurden nach einem modifizierten Protokoll [140, 141] zur Inhibition der Spindelapparate für 2 h bei RT in 2 mM 8-Hydroxyquinolin inkubiert, in eiskaltes Fixativ (Ethanol-Eisessig 3:1) überführt und bei -20°C gelagert. Vor der Präparation wurden die Wurzelspitzen in HPLC-Wasser gewaschen und in 250 µl Enzymmix (2,5 % (w/v) Cellulase, 2,5 % (w/v) Pektinase in 75 mM KCl pH 4,0) 1 h bei 40°C und 10 min bei 55°C inkubiert. Dabei wurde das Material alle 15-20 min mit einem Mikropistill homogenisiert. Die Protoplastenlösung wurde im Anschluss durch eine Nylon-Gaze (Maschenweite 80 µm) gefiltert und mit 1 ml 75 mM KCl 8 min bei RT inkubiert (hypotonische Behandlung). Daraufhin folgten insgesamt vier Zentrifugationsschritte (80 x g, 7 min, RT), wobei das Protoplastenpellet dreimal in 500 µl Fixativ gewaschen und nach der letzten Zentrifugation in 30 µl Fixativ aufgenommen wurde.

2.2.17 CHROMOSOMENPRÄPARATE

Die isolierten Protoplasten (2.2.16) wurden in einer eigens dafür angefertigten Auftropfstation (Abbildung 2.1) auf -20°C vorgekühlte SuperFrost® Plus Objektträger (Menzel-Gläser, Braunschweig, Deutschland) getropft. Nach Abdampfen des Fixativs wurden die Präparate mit einem Tropfen 45 % Essigsäure bedeckt und mit Deckglas versehen etwa 20 min auf Trockeneis inkubiert. Die Deckgläser wurden mit einem Skalpell abgesprengt und die Präparate wurden bis zur Hybridisierung bei -20°C in 100 % Isopropanol gelagert.

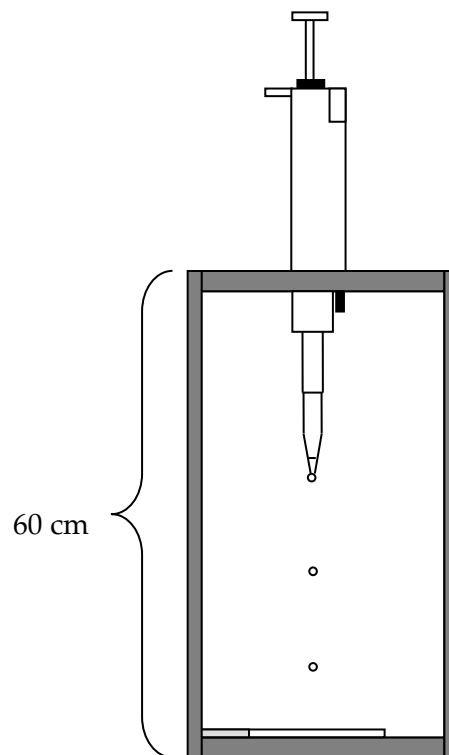


Abbildung 2.1: Auftropfstation. Schemazeichnung der Auftropfhilfe zum Spreiten von Metaphasen bei der Herstellung von Chromosomenpräparaten aus Protoplastenisolationen.

2.2.18 FLUORESCENZ *IN SITU*-HYBRIDISIERUNG (FISH)

2.2.18.1 Vorbereitung der Präparate

Die Präparate wurden als Vorbereitung auf die Hybridisierung in einer a-Hyb™ Hybridisierungsstation (Miltenyi Biotec, Bergisch Gladbach, Deutschland) gemäß folgendem Protokoll behandelt: Alkoholreihe 100 %, 70 %, 50 %, 30 % Isopropanol

(je 5 min 20°C), 0,1 x SSC (5 min 20°C), 2x SSC (30 min 70°C, 5 min 20°C), 0,05 N NaOH (1 min 16°C), 0,1 x SSC (5 min 20°C), Alkoholreihe 30 %, 50 %, 70 %, 100 % Isopropanol (je 5 min 20°C).

2.2.18.2 Sonden

Anhand der Sequenz eines cDNA-Klons aus der Ackerbohne *Vicia faba* (GenBank Eintrag X17535.1) wurden Primer (rRNA_for / rRNA_rev, 2.2.19) für Teile der die 18S und 28S rRNA kodierende rDNA generiert. Das PCR-Produkt (1203 bp) wurde kloniert und per PCR mit Universalprimern wieder amplifiziert. 1 µg dieses Produktes wurde für die Markierung mit dem DIG DNA Labeling Kit (Roche, Indianapolis, USA) eingesetzt. Weiterhin wurden mit dem BioNick™ Labeling System (Invitrogen, Carlsbad, USA) markierte Klone (je 1 µg) einer BAC-Bibliothek (Doktorarbeit Dominik Otto [103]) für die Hybridisierung eingesetzt.

2.2.18.3 Hybridisierung und Detektion

Die Hybridisierung wurde in feuchten Kammern bei 52°C über Nacht durchgeführt. Danach folgten zwei Waschschriffe (RT) in 2 x SSC (5 min) und 1 x PBS (10 min). Die Detektion erfolgte ebenfalls in feuchten Kammern bei 37°C für 3 h, zum Nachweis der DIG-markierten rDNA wurden Fab Fragmente von Anti-Digoxigenin-Fluorescein (30 µg/ml) aus Schaf benutzt, während Biotin-markierte BAC-Klone mit Avidin-Rhodamin (4 µg/ml) detektiert wurden (beides Roche). Nach einem weiteren Waschschriff in 1 x PBS (5 min) wurden die Präparate mit Vectashield® HardSet Mounting Medium mit DAPI (Vector Laboratories, Burlingame, USA) eingedeckelt und bei 4°C ü. N. aushärten lassen. Alle Aufnahmen wurden mit einem BX61 Mikroskop (Olympus, Essex, UK) generiert.

2.2.19 PUFFER UND LÖSUNGEN

10 x E-Puffer

0,36 M TRIS; 0,3 M Natriumdihydrogenphosphat; 0,1 M EDTA in VE-Wasser

10 x MOPS

0,2 M 3-(N-Morpholino)-propansulfonsäure; 0,05 M Natriumacetat; 0,01 M EDTA pH 7,0 in DEPC-Wasser

10 x PBS

32 mM Dinatriumhydrogenphosphat; 5 mM Kaliumdihydrogenphosphat;
13 mM Kaliumchlorid; 135 mM Natriumchlorid pH 7,4 in VE-Wasser

10 x SSC

1,5 M Natriumchlorid; 0,15 M Natriumcitrat in VE-Wasser

10 x TBE

0,89 M TRIS; 0,02 M EDTA; 0,89 M Borsäure in VE-Wasser

Amp/X-Gal/IPTG-haltige Agarplatten

1,5 % (w/v) Agar; 1 % Ampicillin-Stammlösung; 0,1 % X-Gal-Stammlösung;
0,1 % IPTG-Stammlösung in LB-Medium

Ampicillin-Stammlösung

10 mg/ml Ampicillin in HPLC-Wasser

DEPC-Wasser

0,1 % (v/v) Diethylpyrocarbonat (DEPC) in VE-Wasser

DNA-Ladepuffer

30 % (w/v) Saccharose; 0,25 % (w/v) Bromphenolblau in VE-Wasser

Ethidiumbromid-Stammlösung

0,5 % (w/v) Ethidiumbromid in 1x E-Puffer

IPTG-Stammlösung

47,6 mg/ml IPTG in DMF

LB-Medium

2 % (w/v) Trypton; 1 % (w/v) Hefeextrakt; 1 % (w/v) Natriumchlorid in VE-
Wasser

RNA-Färbelösung

50 % (v/v) Glycerin; 1 mM EDTA pH 8,0; 0,25 % (w/v) Bromphenolblau; 0,25 % (w/v) Xylencyanol in VE-Wasser

RNA-Ladepuffer

1 Vol. Formamid; 0,2 Vol. 10x MOPS; 0,3 Vol. Formaldehyd

X-Gal-Stammlösung

100 mg/ml X-Gal in DMF

2.2.20 PRIMER

Name (Gen)	Sequenz (5' – 3')
SP6_universal	TAATACGACTCACTATAGGG
T7_universal	ATTTAGGTGACACTATAG
Md_GAPDH_for	CTGCCACCCAGAAGACTG
Md_GAPDH_rev	GCCCTCAGACTCCTCTTTG
human_GAPDH_for	TGCACCACCAACTGCTTAGC
human_GAPDH_rev	GGCATGGACTGTGGTCATGAG
rRNA_for	CACCGCCCGTCGCTCCT
rRNA_rev	CATCCGCTTCCCTCCCGA
EF1a_for	CGGATATGCCCCAGTTCTG
EF1a_rev	CACGAACAGCAAAGCGACC
βTub_for	GCGTCACCCTCAAATGGTC
βTub_rev	GCTGCTCAAATGCTCGAACG
ADF_for	GAGTTTCTCCACAATCACCTC
ADF_rev	CCTCAGCCTCCACCGTC
LOX2_for	GCTGGATATTTCCCTAACCG
LOX2_rev	CATCAATAATTCCTTCTAGCCTC
AOC4_for	CAAGTCAACACCTCGGCG
AOC4_rev	CAGCTTGAAGGGAAACACTATC

Name (Gen)	Sequenz (5' – 3')
AOS_for	CAACTCAGTTCACACGACTCG
AOS_rev	CTCAAGGGCTGTGATTGAAG
CAX3_for	CATGCAGGAGCAATCATATTTG
CAX3_rev	GAACCAATCCCTTCAGGTAG
PER42_for	CCGTTTCAGTACGTGAGAAATG
PER42_rev	GTGAGAGGGTTGTTCTCGG
SQLE_for	GATTGCTTGAAGTACATGGCTG
SQLE_rev	CTGTACCGCTTCCACTCTGC
BI-1_for	GAGCCTACCTGCATATCCTG
BI-1_rev	CAGAGCAGTTCCTCCAAATGC
EP1_for	CATCTAAGTGTGGTTCGTATG
EP1_rev	GCACTTACAGTCCTTATCACAC
U2af_for	CTCGGTGGTGATTTCTGATAAC
U2af_rev	CATGCTGCCCAATTAGCTTC
MDP0000679810_for	GACAACATGCTAGGAGCAC
MDP0000679810_rev	CGACCAAGTTAAACTTCTAG
MDP0000909874_for	GCAGCAACTGCCTGATGTCT
MDP0000909874_rev	CACACTCTCTTCAGATGGAG
MDP0000146076_for	GATAGATCACCTCCTSGACC
MDP0000146076_rev	CTTGTTTCGGTTGGCTTCCTT
MDP0000565104_for	GCATCATCAAATCTATCACCG
MDP0000565104_rev	CTGGACGCCCTGCTCCGT
MDP0000613462_for	GAACAAGGACATGGAGATGGA
MDP0000613462_rev	CATGATCTTACCAACCACCTC
MDP0000276057_for	ATGCTGGTCTTCCAATTCTTAC
MDP0000276057_rev	CTATCAGATGAGCCACCATTAG
MDP0000247363_for	GAGATGGTCCTTCCCAAGG
MDP0000247363_rev	GACAATATCTTCAGATTCTCC
MDP0000871409_for	GTCATGCTCTTAGCAAGCAAG
MDP0000871409_rev	GCTTTCCTCGCAAGTGGAAG
MDP0000834156_for	CTCTCGCTATGGACTTGCTG

Name (Gen)	Sequenz (5' – 3')
MDP0000834156_rev	CATAGCCATGCAGGATTCTTC
MDP0000504527_for	CGGTCTCTGTGAGGACAGC
MDP0000504527_rev	CACCTTGACGTATCTTGGTTG
MDP0000309906_for	CTGCGTAGTTGTTGACATAATC
MDP0000309906_rev	CTGATCCACTTGGCAGACTG
MDP0000268093_for	GACCCCAACAAGCCCAAGA
MDP0000268093_rev	GTTATATTCCTCTACCTCCTTC
MDP0000192235_for	CTCACATGCTCAAGAAGTGC
MDP0000192235_rev	CATCAAACCCTTGTTGTCCAAG
MDP0000176025_for	CCGCATGTGTTAGCTCAAGC
MDP0000176025_rev	CTTCTCACGCTTTACTGCCAC
MDP0000175733_for	GCAGAAGGATGATCGAACCG
MDP0000175733_rev	CGATATTTGGATTACTCGCTTG
MDP0000175375_for	GAAGAACTACAGGGGAGTGAG
MDP0000175375_rev	CGGAACTCGATTGCAGCTTTG
MDP0000166068_for	CGATCACCACCCTCGTCAC
MDP0000166068_rev	GCAAGAGCACCTTTCTGCT
MDP0000157879_for	GTAATTCCAGACCCAGAGAC
MDP0000157879_rev	CTACCATCACGCTAATGTCTG
MDP0000559039_for	GATGCCTGACGAACAGGTG
MDP0000559039_rev	GTGCTGGGTCCACGTCAAG
MDP0000122958_for	CAACTTGCCCATGTCCGTCG
MDP0000122958_rev	CTGATGACACTCCCATCAC

2.3 NEXT GENERATION SEQUENCING

2.3.1 ROCHE 454

Alle 454-Pyrosequenzierungen (Roche, Indianapolis, USA) wurden mit der GS FLX Titanium Chemie durchgeführt. Ausgehend von je 5 µg ds cDNA (2.2.8.1) wurden die Bibliotheken und Sequenzierungen von der Firma Seq-IT (Kaiserslautern, Deutschland) durchgeführt. Zeitpunkte der Entnahme (Tabelle 2.1) des hierfür eingesetzten Materials waren der 22. Mai 2009 (A14 und P28) und der 29. September 2009 (P28). Es wurde jeweils eine Viertel Picotiterplatte (entspricht ca. 250 k Sequenzen) bestellt. Aufgrund der zum Zeitpunkt der Sequenzierung aktuellen Chemie konnte mit Sequenzlängen von 500 nt gerechnet werden.

2.3.2 ILLUMINA GENOME ANALYZER IIX

Die Illumina-Sequenzierung auf dem Genome Analyzer Iix (Illumina, San Diego, USA) und vorherige Erstellung der Bibliotheken wurden vom Cologne Center for Genomics (CCG, Köln, Deutschland) durchgeführt. Dazu wurden jeweils 10 µg aRNA (2.2.5) von Material der Entnahmezeitpunkte 22. Mai 2009 (A14) und 29. September 2009 (P28) verwendet. Angefordert wurde hier je eine Lane (für A14 und P28) eines 95 bp Paired-End-Laufes.

2.3.3 ILLUMINA HISEQ 2000

Zwei weitere Sequenzierungen wurden auf dem hauseigenen HiSeq2000 (Illumina, San Diego, USA) durchgeführt. Die Erstellung der Bibliotheken wurde von Mitarbeitern des Instituts für Molekulargenetik (IMSB, JGU Mainz, Deutschland) durchgeführt. Als Ausgangsmaterial diente zum einen ds cDNA von Material des Entnahmezeitpunktes 22. Mai 2009 (A14 und P28) sowie aRNA von Material des Entnahmezeitpunktes 20. Juli 2010 (A14 und P28). Insgesamt wurden zwei Lanes einer Flowcell sequenziert, wobei die Bibliotheken eines Entnahmezeitpunktes zu gleichen Anteilen auf jeweils eine Lane aufgeteilt wurden (Multiplex).

2.4 BIOINFORMATISCHE METHODEN

2.4.1 KONVERTIERUNG DER DATENFORMATE

Für die verschiedenen Folgeanwendungen mussten die Rohdaten zunächst in ein passendes Datenformat konvertiert werden.

Die Roche 454-Sequenzdaten wurden in SFF-Dateien (Standard Flowgram Format) von der Firma Seq-IT geliefert. Diese wurden mit Hilfe des Perl-Skriptes *roche454ace2caf* (<http://genome.imb-jena.de/software/roche454ace2caf/>) in das leichter handhabbare FASTA Format (*.fna) und die dazugehörige Qualitätsinformationsdatei (*.qual) konvertiert. Etwaige Adaptersequenzen oder Barcodes wurden im Anschluss aus der FASTA-Datei entfernt.

Die mit dem Genome Analyzer Iix sequenzierten cDNA-Bibliotheken wurden im QSEQ Format vom Cologne Center for Genomics bereitgestellt. Diese wurden mit dem Perl-Skript *qseq2fastq* [Supporting Information (SI) Perl-Skript_1] in FASTQ-Dateien (Sequenzen mit zugehörigen Qualitätswerten im ASCII-Code) beziehungsweise mit dem Perl-Skript *qseq2fasta* [SI Perl-Skript_2] direkt in FASTA-Dateien (ohne Qualitätswerte) umgewandelt. HiSeq2000 Sequenzdaten wurden ebenfalls im QSEQ Format bereitgestellt und mit den gleichen Skripten in die oben genannten Formate umgewandelt. Bei all diesen Formatierungen wurde ein erster Filter (in den Skripten enthalten) angewendet, der die Sequenzen nach ihrem Abschneiden im Standard Illumina Qualitätsfilter (chastity) separiert. Der chastity-Wert (2.1) ist definiert als der Quotient aus der höchsten Basenintensität (I_{\max}) und der Summe der höchsten beiden ($I_{\max}, I_{2\max}$).

$$(2.1) \quad \text{chastity} = \frac{I_{\max}(A, G, C, T)}{I_{\max}(A, G, C, T) + I_{2\max}(A, G, C, T)}$$

Auf diese Art und Weise lassen sich Cluster mit einem niedrigen Signal-Rausch-Verhältnis identifizieren. Wird während der Sequenzierung der Schwellenwert von 0,6 zweimal überschritten, wird die Sequenz mit 0 bewertet, ansonsten mit 1. Nach diesem Kriterium, das beim Auslesen im Sequenznamen integriert wird, kann das Skript die Sequenzen separieren.

2.4.2 ANALYSE DER ROHDATEN

Alle Rohdaten (Roche 454 und Illumina) wurden nach der Konvertierung ins FASTA-Format statistisch mit dem Perl-Skript *pasta* [SI Perl-Skript_3] ausgewertet. Dabei wurde die Anzahl und durchschnittliche Länge der Sequenzen, der N50- und N80-Wert sowie die Basenzusammensetzung ausgelesen.

2.4.3 ASSEMBLIERUNG DER ROHDATEN

2.4.3.1 *Assemblierung der Roche 454-Rohdaten*

Das *de novo* Assemblierung der Roche 454-Rohdaten wurde mit SeqMan NGen (DNASTAR, Madison, USA) auf Grundlage der von Seq-IT gelieferten SFF (P28) bzw. der in FASTA konvertierten SFF Datei (A14) durchgeführt. Dabei wurden folgende Parameter verwendet: match size = 19; match spacing = 10; minimum match percentage = 85; match score = 10; mismatch penalty = 25; gap penalty = 25; max gap = 15; expected coverage = 500.

2.4.3.2 *Assemblierung der Illumina-Rohdaten*

Aufgrund der extrem großen Datenmenge wurde beim *de novo* Assemblierung der Illumina-Rohdaten auf das Programm ABySS (Assembly By Short Sequences) zurückgegriffen [142, 143]. Der zu Grunde liegende Algorithmus beruht auf der verteilten Repräsentation (engl. Distributed representation) eines de Bruijn-Graphen, welches eine hohe Parallelisierung ermöglicht [144, 145]. Hierzu werden die Einzelsequenzen zunächst in k-mere zerlegt, die die Knoten des de Bruijn-Graphen repräsentieren. Bei einer Überlappung von k-1 werden zwei Knoten über eine Kante miteinander verbunden und die entsprechenden Sequenzen werden zusammengefügt. Im nächsten Schritt werden die Paired-End-Informationen dazu genutzt, die bestehenden Contigs zu verifizieren und zu verlängern. Die unformatierten QSEQ-Daten wurden mit folgenden Parametern assembliert: k-mer-Länge k = 83; Paired-End-Daten n = 2; Coverage c = 10; End Trimming e = 5; Seed Length s = 2k = 166.

2.4.3.3 Assemblierung kleinerer Datenmengen

Bei Assemblierungen kleiner Datenmengen, beispielsweise von kartierten Rohdaten gegen eine geeignete Referenz (2.4.5), wurde auf das Programm SeqMan (DNASTAR, Madison, USA) zurückgegriffen. Dieses Programm verwendet einen modifizierten Martinez-Needleman-Wunsch-Algorithmus [146, 147] für die Generierung lokaler Alignments, die dann in Contigs zusammengefasst werden.

2.4.4 BLAST

Zum Abgleich der Sequenzdaten mit verschiedenen Datenbanken wurde der BLAST (Basic Local Alignment Search Tool)-Algorithmus zur Erstellung von lokalen Sequenzalignments eingesetzt [148], welcher auf einer heuristischen Annäherung des Smith-Waterman-Algorithmus [149] basiert. Hierbei werden die Suchsequenzen (query) in kleine words (Standardlänge 11 bp) zerlegt und mit einer Datenbank im Hinblick auf das beste lokale Alignment abgeglichen. Im Falle eines Treffers werden die words nach rechts und links unter Einbeziehung von Lücken verlängert, bis die Sequenzen keine Übereinstimmung mehr zeigen. Die Qualität des auf diese Weise gefundenen Alignments wird mit Hilfe von zwei Werten, dem Score (Ähnlichkeit der beiden Sequenzen) und dem Erwartungswert (E-Wert, ein Maß für die Wahrscheinlichkeit, ein zufälliges Alignment mit dem gleichen Score zu finden) angegeben. Die Qualität des Alignments steigt mit zunehmendem Score und abnehmendem E-Wert.

Der größte Teil der BLAST-Suchen wurde parallelisiert auf dem HPC-Cluster (High Performance Computing Cluster) der Universität Mainz durchgeführt. Die Suchen wurden dabei mit dem Perl-basierten Wrapper *Blaster1.2* [SI Perl-Skript_5] initiiert. BLAST-Suchen einzelner DNA- oder Proteinsequenzen wurden meist direkt gegen das sequenzierte Apfelgenom [9] (<http://genomics.research.iasma.it/gb2/gbrowse/apple/> bzw. http://www.rosaceae.org/gb/gbrowse/malus_x_domestica/) oder die gut annotierte UniProtKB/Swiss-Prot Proteindatenbank [150] von der Seite des National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>) durchgeführt. Eigene Datenbanken wurden aus geeigneten FASTA-Dateien mit *formatdb* (NCBI) erstellt, die BLAST-Suchen dagegen wiederum auf dem HPC-Cluster oder lokal mit der Hilfe von *blastall* oder *blast client* (NCBI).

2.4.5 VERARBEITUNG DER BLAST-AUSGABE MIT BORT

Die umfangreichen BLAST-Ausgaben enthalten Namen- und Längeninformationen der Suchsequenz (query) und der Zielsequenz (description), den Score- und E-Wert, die Anzahl der Übereinstimmungen (identities), die Orientierung sowie das Alignment selbst. Um die Ausgaben multipler BLAST-Suchen mit bis zu 80.000.000 Einträgen übersichtlich und handhabbar zu machen, wurde das Perl-Skript BORT (BLAST Output Refinement Tool) etabliert [SI Perl-Skript_6]. BORT parst die Sequenzalignments und sortiert diejenigen ohne Treffer (hits) in eine NoHits-Datei aus, um die Datenmenge zu verringern. Weiterhin werden alle Informationen der Sequenzen mit Treffer auf eine einzige Textzeile komprimiert und in einer sogenannten BORT-Datei zwischengespeichert. Aus dieser Datei kann dann eine individuelle Ausgabedatei erstellt werden, in der alle relevanten Informationen (z.B. query, description und E-Wert) unter Einhaltung der Filterkriterien (z.B. E-Wert $\leq 1E^{-10}$) aufgeführt werden. Des Weiteren lässt sich die BLAST-Ausgabe auch quantifizieren, was die Grundlage für die Expressionsanalyse darstellt. Dazu werden alle „descriptions“ kollabiert und die Anzahl der „reads“ bzw. Contigs (mit der Anzahl an „reads“, aus denen diese assembliert wurden) wird aufaddiert. Diese quantitativen Informationen können exportiert und in EXCEL weiterverarbeitet werden.

2.4.6 STATISTISCHE ANALYSE DER DIFFERENTIELLEN REGULATION

Um aus der reinen Anzahl der „reads“ für ein bestimmtes Gen ein statistisch abgesichertes Maß für die Signifikanz der differentiellen Genexpression zu erhalten, wurde die für cDNA-Banken entwickelte Methode von Stekel et al. (2000) auf die Daten von A14 und P28 angewendet [151]. Die Signifikanz wird durch den R_j -Wert definiert, die Berechnung beruht auf der Log-Likelihood-Ratio-Statistik für ein Gen j , die sich asymptotisch einer χ^2 -Verteilung nähert.

$$(2.2) \quad R_j = \sum_{i=1}^m x_{i,j} \log \left(\frac{x_{i,j}}{N_i f_j} \right)$$

Dies beschreibt Formel 2.2, mit m als Anzahl der zu Grunde liegenden cDNA-Banken bzw. Datensätze, $x_{i,j}$ als Anzahl der „reads“ des Gens j in Datensatz i und N_i als die Anzahl aller Sequenzen in Datensatz i . f_j ist ein Maß für die Häufigkeit eines Gens j in sämtlichen Datensätzen und wird mit Hilfe von Formel 2.3 berechnet.

$$(2.3) \quad f_j = \frac{\sum_{i=1}^m x_{i,j}}{\sum_{i=1}^m N_i}$$

Statistisch signifikant reguliert sind Gene mit einem R_j -Wert ≥ 8 , wobei das Maß der Signifikanz mit steigenden R_j -Werten zunimmt.

2.4.6.1 RPKM-Normalisierung

Ebenfalls zur Analyse der differentiellen Regulation und zur Identifizierung der häufigsten Transkripte wurde die Anzahl der „reads“ pro Referenzgen RPKM-normalisiert [152]. Dabei muss die Länge der Transkripte bekannt sein, um die Zahl der Reads pro Kilobase Transkript pro Millionen kartierter „reads“ zu erhalten. Der RPKM-Wert lässt sich mit Hilfe von Formel 2.4 berechnen.

$$(2.4) \quad \text{RPKM} = \frac{\text{Reads}_{\text{kartiert}} \times 10^6}{\text{Transkript (kb)} \times \text{Reads}_{\text{tot}}}$$

2.4.7 KARTIERUNG VON ROHDATEN

2.4.7.1 PASS

Für das Kartieren (Mapping) kurzer Sequenzen (wie sie beim Next Generation Sequencing generiert werden) gegen eine Referenzsequenz wurde das speziell dafür entwickelte Programm PASS [153] eingesetzt. Dabei wird in einem ersten Schritt aus der Referenzsequenz ein Index von seed words (Standardlänge 12 bp) und ein precomputed score table (PST) aller Alignments der kurzen Sequenzen gegen einander erstellt, der im RAM gespeichert wird. Dieser Schritt stellt die Grundlage für die hohe Geschwindigkeit dar, mit der PASS arbeitet. Im zweiten Schritt werden die zu kartierenden Sequenzen mit dieser Information abgeglichen, wobei zunächst seed words im Referenzindex identifiziert, die flankierenden Regionen mittels PST

analysiert und die aufgearbeiteten Alignments in einer GFF-Datei ausgegeben werden.

Illumina-Daten wurden im FASTQ Format eingeladen und mit PASS gegen Referenzen im FASTA-Format kartiert. Als Parameter dienten eine Sequenzidentität von 90 %, eine minimale Qualität der Eingangssequenzen von 30 (bei einer Skala von 0 – 75) auf einer Länge von 60 Basen und nicht mehr als drei Lücken bzw. Mismatches am Ende einer Sequenz. Bei gespleißten Alignments (also die Zerlegung Intron-überspannender Sequenzen bei genomischen Referenzen) wurden eine maximale Intronlänge von 50.000, eine Überlappung von 10 Basen und eine minimale Gesamtlänge von 95 Basen für beide Teilsequenzen gewählt. Die so kartierten Sequenzen wurden mit dem Perl-Skript *poet* [SI Perl-Skript_4] aus der GFF-Ausgabedatei extrahiert.

454-Daten wurden in SFF- bzw. FASTA-Dateien mit den gleichen Parametern kartiert, allerdings wurde hierbei eine minimale Qualität von 20 (auf einer Skala von 0 – 40) vorausgesetzt.

2.4.7.2 *Bowtie*

Sehr stringente Kartierungen wurden mit dem Programm Bowtie [154, 155] durchgeführt. Bowtie rechnet mit einer Burrows-Wheeler-Transformation (BWT) basierend auf einem full-text minute-space (FM) Index. Die BWT ist eine reversible Permutation der Charaktere in einem Text und erlaubt es, große Texte mit vergleichsweise wenig RAM zu durchsuchen, wodurch Bowtie auf herkömmlichen Desktop-PCs mit extrem hohen Geschwindigkeiten arbeiten kann. Die Suche in einem FM Index wird gewöhnlich mit einem Algorithmus nach Farragina und Manzini [156] realisiert, dieser ist in Bowtie jedoch soweit modifiziert, dass bis zu drei Mismatches zugelassen werden können.

Die Indizes wurden mit der im Bowtie-Wrapper *Bowtier1.2* [SI Perl-Skript_7] integrierten Funktion ‚*bowtie-build*‘ mit den Parametern `-a -o 3 -t 12` erstellt. Die Illumina-Daten wurden im FASTA-Format eingeladen und mit den folgenden Parametern ebenfalls in *Bowtier1.2* kartiert: `-v 3 -chunkmbs 128 -p 8 -best`. Die Verarbeitung der Ausgabedatei erfolgte analog zum Verfahren mit BORT (2.4.5) mit dem Perl-basierten Skript *MORT* (*Mapping Output Refinement Tool*) [SI Perl-Skript_8].

2.4.8 FUNKTIONELLE ANALYSE MIT MAPMAN

Für die funktionelle Analyse der differentiell regulierten Gene wurde das für pflanzliche Transkriptom- bzw. Proteomdaten entwickelte Programm MapMan verwendet [157, 158]. Gene bzw. Proteine werden darin aufgrund ihrer Annotation in funktionelle Gruppen und Untergruppen eingeteilt (BINs und subBINs). Gegen diese Kartierung können Datensätze abgeglichen und jedes Gen dann einzeln visualisiert werden. Bei komparativen Expressionsanalysen kann der Unterschied in der Expressionsrate zudem farblich hervorgehoben werden, wobei die Farbkodierung abhängig vom Zahlenwert ist.

Da keine geeignete Kartierung für *Malus x domestica* existiert, wurde auf die im Zuge des Pappel (*Populus trichocarpa*)-Genomprojektes [159] annotierten und kartierten Gene zurückgegriffen. Dazu wurden die im Apfel-Genomprojekt [9] annotierten Gene per BLASTp-Suche mit denen der Pappel abgeglichen. Die „read counts“ der am stärksten differentiell regulierten Gene wurden auf 1 in der Referenz A14 normiert und beide Werte (A14, P28) wurden logarithmiert. Diese Werte mit der dazugehörigen Geninformation wurden als XLS-Datei in MapMan importiert.

2.4.9 ANALYSE DER DIFFERENTIELLEN EXPRESSION MIT DESEQ

Die Analyse der differentiellen Expression in allen vorliegenden Datensätzen wurde mit DESeq (R/Bioconductor Paket) durchgeführt [160]. Im Gegensatz zu den meisten Tools, die differentielle Genexpression mit Poisson-Verteilungen testen [161, 162], basiert DESeq auf einer negativen Binomialverteilung, wobei Varianz und Mittelwert durch lokale Regression verknüpft sind. Für die Analyse wurden alle Datensätze mit der gleichen Referenz abgeglichen (BLAST, Kartierung), die „read counts“ wurden normalisiert und in DESeq eingeladen. Dabei wurden die drei Datensätze von A14 als Referenz, die drei Datensätze von P28 als Probe definiert. Die Analyse wurde mit Standardparametern durchgeführt.

3 ERGEBNISSE

3.1 EXTRAKTION UND AUFARBEITUNG DER RNA

Basierend auf der Annahme, dass die Analyse des Transkriptoms von Sprossapikalmeristemgewebe am wahrscheinlichsten Aufschluss über den kolumnaren Phänotyp geben dürfte, wurden die in Geisenheim gesammelten Triebspitzen vor der RNA-Extraktion gezielt präpariert. Dabei wurde das Meristemgewebe weitgehend von Spross- und Blattgewebe befreit und junge Blattanlagen vollständig entfernt. Vor dem Hintergrund, dass mechanische Verwundungen wie diese eine schnelle und systematische Reaktion auf transkriptioneller Ebene hervorrufen [163, 164], wurde nach Herbeiführen des abiotischen Stimulus das Gewebe schnellstmöglich mit flüssigem Stickstoff tiefgekühlt und mit der RNA-Extraktion begonnen. Das Ergebnis einer solchen Präparation zeigt Abbildung 3.1 A und B. Diese Elektropherogramme (Agilent Bioanalyzer) geben Aufschluss über Quantität und Qualität (RIN)¹ der Nukleinsäure. Hier sind Total-RNAs von A14 (A) und P28 (B) mit sehr guten RINs von 9,6 bzw. 9,5 (auf einer Skala von 0 bis 10) gezeigt. Deutlich überrepräsentiert sind die 5,8S (ca. 180 nt), 18S (ca. 1800 nt) und die 28S rRNA (ca. 3300 nt), da ribosomale RNAs etwa 90-95 % der gesamten RNA einer Zelle ausmachen. Vor dem Peak der 18S rRNA treten zudem noch weitere kleinere Peaks auf, welche plastidäre und mitochondriale rRNA repräsentieren. Da nicht alle extrahierten Total-RNAs eine solch hohe Qualität wie die hier gezeigten aufwiesen und die spätere Poly-A⁺-Aufreinigung eine große RNA-Menge voraussetzte, wurden nur RNAs mit einem RIN von bis zu 6,0 weiterverwendet. Für die folgende Aufreinigung der mRNA wurden mindestens 100 µg Total-RNA aus insgesamt bis zu 16 Extraktionen eingesetzt, wobei das zu Grunde liegende Meristemgewebe stets von mehreren

¹ RIN (RNA integrity number) ist ein Algorithmus für die Bestimmung der Qualität (Integrität) von RNA. Zur Berechnung wird eine Kombination verschiedener Laufeigenschaften nach elektrophoretischer Auftrennung verwendet.

Bäumen (Klonen) stammte. Abbildung 3.1 C und D zeigen die mRNAs von A14 (C) und P28 (D) nach Anreicherung mit oligo(dT) Beads. Nach der Anreicherung von Poly-A⁺-RNA betrug der Anteil der rRNA noch 9,7 % bzw. 6,5 %.

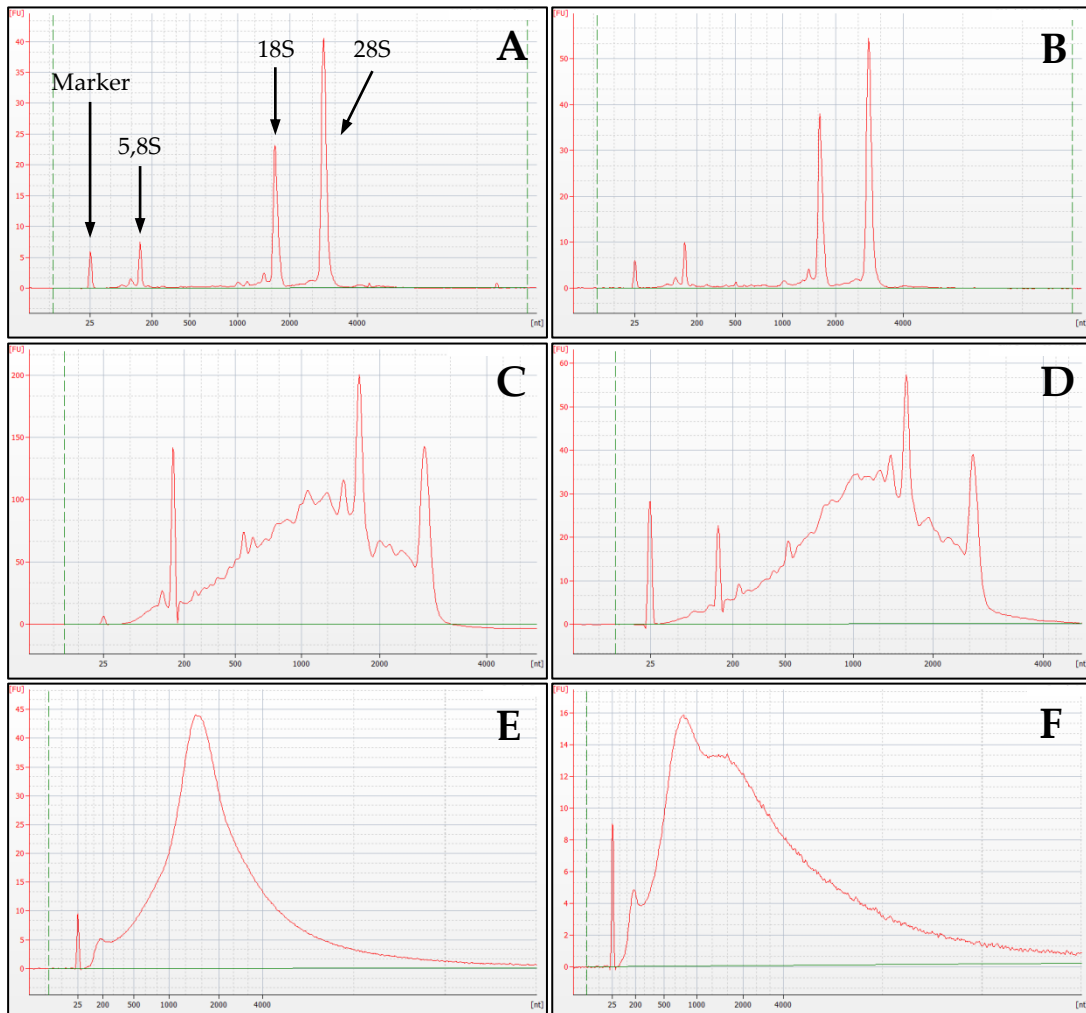


Abbildung 3.1: Probenaufbereitung. Abgebildet sind Elektropherogramme (Agilent Bioanalyzer) der verschiedenen RNAs. Der Peak bei 25 nt repräsentiert den unteren Marker (A-F, laufintern). A und B Gesamt-RNA aus Sprossapikalmeristemen von A14 (A) und P28 (B). Die überrepräsentierten Peaks der 5,8S, 18S und 28S rRNA sind markiert. C und D mRNA nach einer Runde Aufreinigung mit oligo(dT) Latex Beads. Der Grad der rRNA-Kontamination liegt bei 9,7 % (A14, C) bzw. 6,5 % (P28, D). E und F Amplifizierte aRNA von A14 (E) und P28 (F).

Da die Konzentration der mRNA aber mit durchschnittlich 30 ng/μl für die Synthese von ds cDNA bei Weitem nicht ausreichend war, musste auf eine lineare Amplifikation zurückgegriffen werden. Zu späteren Zeitpunkten dieser Arbeit

erlaubte die aktuelle Sequenzierchemie zwar den Einsatz von weitaus weniger Material, aus Gründen der Vergleichbarkeit wurden jedoch alle Proben identisch behandelt. Das Ergebnis der Amplifikation zeigt Abbildung 3.1 E und F. Es ist keine rRNA-Kontamination mehr zu erkennen, da in der Amplifikation ein zweiter Poly-A⁺-Aufreinigungsschritt inbegriffen ist. Lediglich im Bereich des 5,8S Peaks ist auch in der aRNA ein lokales Maximum zu erkennen. Wie auch in der mRNA erstreckt sich der Größenbereich von knapp über 25 nt bis 4000 nt, wobei der größte Anteil zwischen 1300 und 2000 nt (A14, E) bzw. 750 und 2000 nt (P28, F) liegt. Auffällig ist jedoch, dass bei beiden amplifizierten Proben im Gegensatz zur mRNA (C und D) die Kurven jenseits von 4000 nt nicht mehr die Basislinie (grün) berühren, sondern sich im Bereich höherer Molekulargewichte dieser nur annähern. Dies kann auf Artefakte der Amplifikation zurückgeführt werden, da hierbei durch ungewollte Hybridisierung von RNAs durchaus scheinbar längere Fragmente entstehen können.

3.2 SEQUENZIERUNGEN UND ANALYSE DER ROHDATEN

Im Verlauf dieser Arbeit wurden aus den in 3.1 beschriebenen aRNAs insgesamt neun Sequenzierungen durchgeführt, drei davon mit der Roche 454-Technologie, zwei mit dem Illumina GA_{ITX} und vier mit dem Illumina HiSeq2000. Alle Sequenzen sind im EBI Sequence Read Archive (SRA) unter ERP000629 abgelegt. Die notwendige Synthese doppelsträngiger cDNA (ds cDNA) wurde dabei für die Roche 454-Sequenzierungen selbst durchgeführt, während dies im Falle der Illumina-Sequenzierungen als Teil der Bibliothekerstellung vom jeweiligen Sequenzierservice übernommen wurde. Die Ergebnisse aller Sequenzierungen sind in Tabelle 3.1 zusammengefasst. Den derzeit existierenden 335.912 ESTs (Expressed Sequence Tags) für die Gattung *Malus* (NCBI *Malus* EST Datenbank, Stand 07.10.2011) konnten somit weitere 489.827.756 Sequenzen hinzugefügt werden, wobei diese Zahl noch etwa 10 % Sequenzen mit geringer Qualität (nicht identifizierbare Basen (N), niedrige Komplexität) beinhaltet. Für die im Folgenden gezeigte Analyse der Rohdaten werden vier Datensätze (je zwei von A14 und P28) analysiert, die auf der gleichen Ausgangs-aRNA basieren und sowohl per Roche 454- als auch mit der Illumina-Technologie sequenziert wurden (siehe Tabelle 3.1).

Tabelle 3.1: Übersicht aller Transkriptomsequenzierungen. Die unten stehende Tabelle gibt eine detaillierte Übersicht über alle Sequenzierungen (getrennt nach Zeitpunkt der Materialentnahme). Die Anzahl der Sequenzen, mit denen alle beschriebenen Folgeanwendungen durchgeführt wurden, sind für A14 und P28 einzeln und in der Summe aufgeführt. Die für die spätere Assemblierung (3.3) eingesetzten Datensätze sind kursiv (SeqMan NGen) bzw. unterstrichen (ABYSS) dargestellt.

Zeitpunkt der Materialentnahme	Art der Sequenzierung	Sequenzen A14	Sequenzen P28
22. Mai 2009	Roche 454	276.038	32.609
	Illumina GA _{fix}	<u>80.289.084</u>	-
	Illumina HiSeq	44.134.596	67.515.702
29. Sept. 2009	Roche 454	-	245.297
	Illumina GA _{fix}	-	<u>76.177.216</u>
20. Juli 2010	Illumina HiSeq	58.616.798	162.540.416
		183.316.516	306.511.240

Die Roche 454-Daten von A14 und P28 (Tabelle 3.2) zeigen eine durchschnittliche Sequenzlänge von 398 (A14) bzw. 401 Basen (P28), eine viertel Picotiterplatte (276.038 bzw. 245.297 Sequenzen) liefert somit 109,9 bzw. 98,5 Megabasen (Mb) an Sequenzinformation. Die N80- bzw. N50-Werte beider Bibliotheken sind mit 484 bzw. 516 Basen identisch, auch die nicht identifizierten Basen stellen jeweils einen Anteil von 0,03 % dar. Lediglich der G/C-Gehalt weicht mit 0,3 % (45,1 % in A14 bzw. 45,4 % in P28) leicht ab.

Tabelle 3.2: Roche 454-Rohdaten (Statistik). Ergebnisse der statistischen Auswertung der Roche 454-Rohdaten von A14 und P28 mit dem Perl-Skript *pasta*.

	A14	P28
Gesamtanzahl Sequenzen	276.038	245.297
Gesamtanzahl Basen	109.909.375	98.471.296
Ø Sequenzlänge (Basen)	398	401
N80 (Basen)	484	484
N50 (Basen)	516	516
G/C-Gehalt (%)	45,1	45,4
Nicht identifizierte Basen (%)	0,03	0,03

Mit rund acht Gigabasen (Gb) liefern die A14 und P28 Illumina-Daten des selben Entnahmezeitpunkts (Tabelle 3.3) jeweils einer Spur die etwa 80-fache Datenmenge verglichen mit den Roche 454-Daten (Unterstrichen und kursiv dargestellte Zahlen in Tabelle 3.1). Da die Sequenzlänge nur 100 Basen beträgt (ebenfalls Ø, N80 und N50), entspricht die Datenmenge 80.289.084 bzw. 76.177.216 Sequenzen („reads“). Die offensichtliche Fehlerquote, ausgedrückt durch die Zahl der mit N bezeichneten Basen², ist mit 0,06 % jedoch doppelt so hoch wie die Fehlerquote bei den 454-Daten. Der G/C-Gehalt weicht mit 2,7 % im Vergleich von A14 (47,2 %) und P28 (44,5 %) wesentlich deutlicher ab als dies bei den Roche 454-Daten der Fall ist.

Tabelle 3.3: Illumina-Rohdaten (Statistik). Ergebnisse der statistischen Auswertung der Illumina-Rohdaten von A14 und P28 mit dem Perl-Skript *pasta*.

	A14	P28
Gesamtanzahl Sequenzen	80.289.084	76.177.216
Gesamtanzahl Basen	8.028.908.497	7.617.721.563
Ø Sequenzlänge (Basen)	100	100
N80 (Basen)	100	100
N50 (Basen)	100	100
G/C-Gehalt (%)	47,2	44,5
Mit N gecallte Basen (%)	0,06	0,06

Abbildung 3.2 zeigt die Verteilung der Leseweiten der Roche 454-Daten nach Entfernen von technischen Sequenzen³. Die Maxima liegen mit 496 (A14) bzw. 488 Basen (P28) dicht am Erwartungswert von 500 Basen (Stand der Sequenzierchemie: März 2010). Nach dem Maximum fallen beide Kurven rasch ab, die „reads“ mit Längen von über 500 Basen zeigen zudem häufig ausgedehnte Homopolymerruns oder eine niedrige Komplexität (hoher A/T-Gehalt).

² Im Zuge einer Sequenzierung wird dem gemessenen Signal (z.B. Fluoreszenz) automatisch die jeweilige Base (fluoreszenzgekoppelt) zugeordnet („base calling“). Kann dies nicht eindeutig erfolgen (z.B. zu hoher Hintergrund, überlagernde Signale), so wird dieser Position in der Sequenz ein „N“ zugeordnet (nicht identifizierbar).

³ Die Entfernung technischer Sequenzen („trimming“) ist notwendig, um lauffintern und anwendungsspezifisch verwendete Sequenzen wie Barcodes (Multiplex) oder Adapter (Erstellung der Bibliothek) nicht mit in die Analysen zu übernehmen.

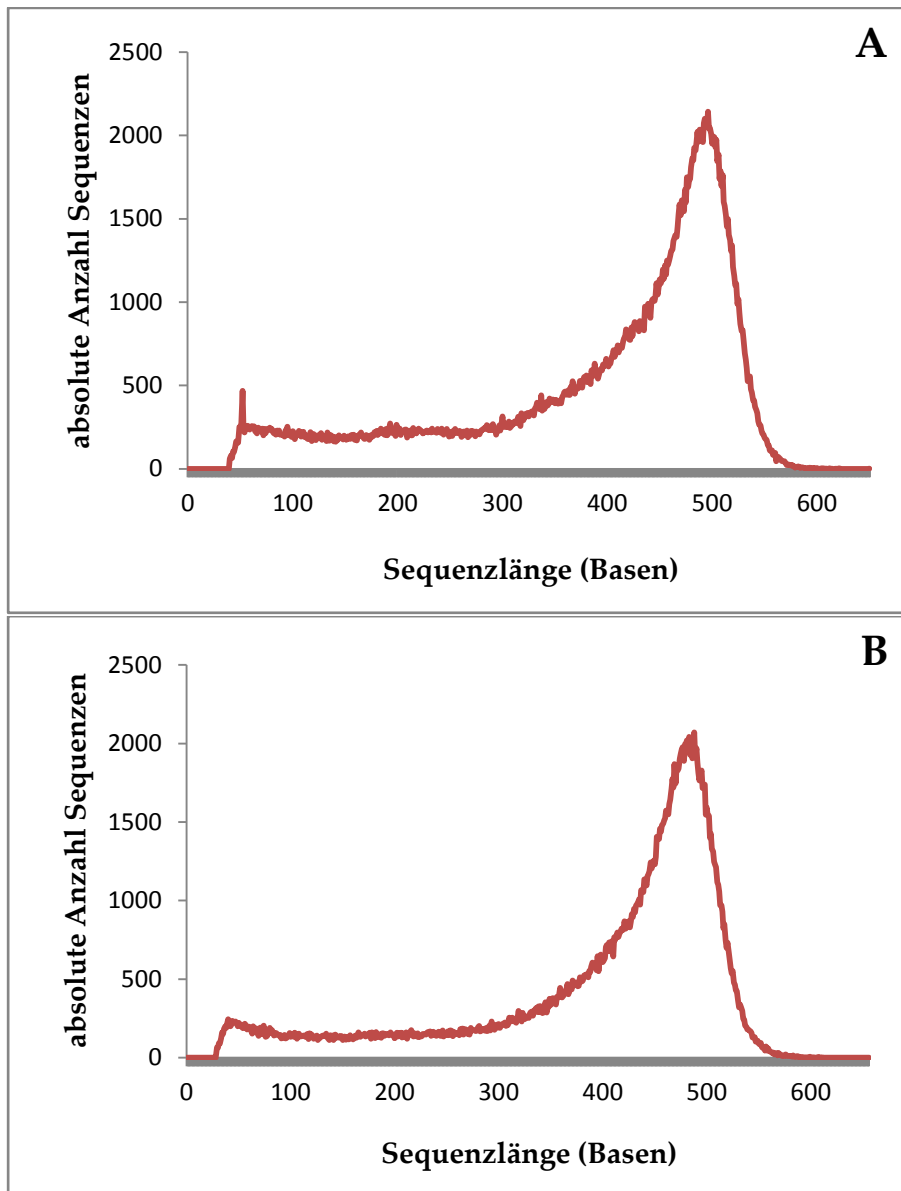


Abbildung 3.2: Roche 454-Rohdaten. Aufgeführt ist die Anzahl der Sequenzen für A14 (A) und P28 (B) gegen die jeweilige Leselänge, wobei die Mittelwerte mit Werten von 398 (A14) bzw. 401 Basen (P28) sehr gut übereinstimmen. In beiden Fällen liegt das Maximum mit 496 (A14) und 488 (P28) nahe an den erwarteten 500 Basen.

Zum weiteren Vergleich der Rohdaten wurden alle Illumina-Sequenzen gegen die assemblierten Contigs (vgl. 3.3) und Singletons der Roche 454-Daten mit PASS [153] kartiert. Bei einer minimalen Identität von 90 % lassen sich so 82 % der A14 Illumina-Sequenzen und 76 % der P28 Illumina-Sequenzen gegen ihre jeweilige Referenz kartieren.

Anhand von fünf ausgewählten mRNA Sequenzen (NCBI *Malus x domestica* Unigene Datenbank) wurde die Abdeckung durch die unterschiedlichen Rohdaten ermittelt, wobei auch hier gegen die jeweilige Referenz mit PASS kartiert wurde und die erfolgreich kartierten Sequenzen anhand der Referenz in SeqMan assembliert wurden. Die Ergebnisse sind in Abbildung 3.3 dargestellt. Dabei wird ersichtlich, dass die Sequenzen überproportional stark nahe dem 3'-Ende kartieren (Abbildung 3.3 E). Dies ist sehr wahrscheinlich auf die zweimalige Selektion mittels oligo(dT)-Affinitätsaufreinigung (Poly-A⁺-RNA und anschließende Amplifikation) zurückzuführen. Die Abdeckung der Sequenzen am 3'-Ende und im 3'-UTR (nicht gezeigt) ist bei den hier gezeigten Genen bis zu 10000-fach, nimmt aber in Richtung 5'-Ende schnell ab. Nur zwei der insgesamt fünf untersuchten mRNA Sequenzen sind von beiden Datensätzen vollständig, d.h. inklusive des 5'-Endes, abgedeckt (WRKY70, 912 bp und Spermidin-Synthase, 1005 bp, Abbildung 3.3 A und B). Diese beiden mRNAs sind auch die kürzesten der betrachteten Sequenzen. Bei längeren Transkripten zeichnet sich eine bessere Abdeckung durch die Illumina-Sequenzen im 5'-Bereich ab; der Zugewinn an Sequenzinformation (bezogen auf die Abdeckung der Referenz) im Vergleich zu den Roche 454-Daten beträgt für AOS (1653 bp, Abbildung 3.3 C) 6 %, für COI1 (1760 bp, Abbildung 3.3 E) 23 % und für MYC2 (2280 bp, Abbildung 3.3 D) 38 %. Die Möglichkeit, dass Lücken in der Abdeckung (Abbildung 3.3 C und E) auf nicht korrekt annotierte Introns bzw. alternatives Spleißen zurückgehen, konnte durch die Bestimmung pflanzentypischer Konsensus-Spleißstellen weitgehend ausgeschlossen werden. Mit den hier durchgeführten Assemblierungen konnte zudem der Grad der Überlappung der Illumina-Sequenzen (da es sich um „Paired-End“-Läufe⁴ handelt) ermittelt werden. Die Überlappung beträgt bei Sequenzpaaren der A14 Bibliothek durchschnittlich 43 Basen, während bei P28 durchschnittlich 47 Basen überlappen (Tabelle 3.4, Seite 52). Dies entspricht einer Redundanz von 21,5 bzw. 23,5 %, was jedoch in gleichem Maße die Genauigkeit durch zweifache Sequenzierung steigert.

⁴ Bei der Generierung von Illumina-Bibliotheken werden am 3'- und 5'-Ende der zu sequenzierenden DNAs unterschiedliche Adapter ligiert (A und B). Nach Amplifikation dieser Fragmente zu sog. Clustern kann sowohl mit einem zu A sowie zu B komplementären Sequenzierprimer die Synthese (= Sequenzierung) gestartet werden. Im Anschluss an diese erste Sequenzierung („single read“) können die Cluster regeneriert und die Synthese mit dem im ersten Durchgang nicht verwendeten Sequenzierprimer gestartet werden („paired end“). Weitere Details zur Sequenzierchemie können Abschnitt 1.4 entnommen werden.

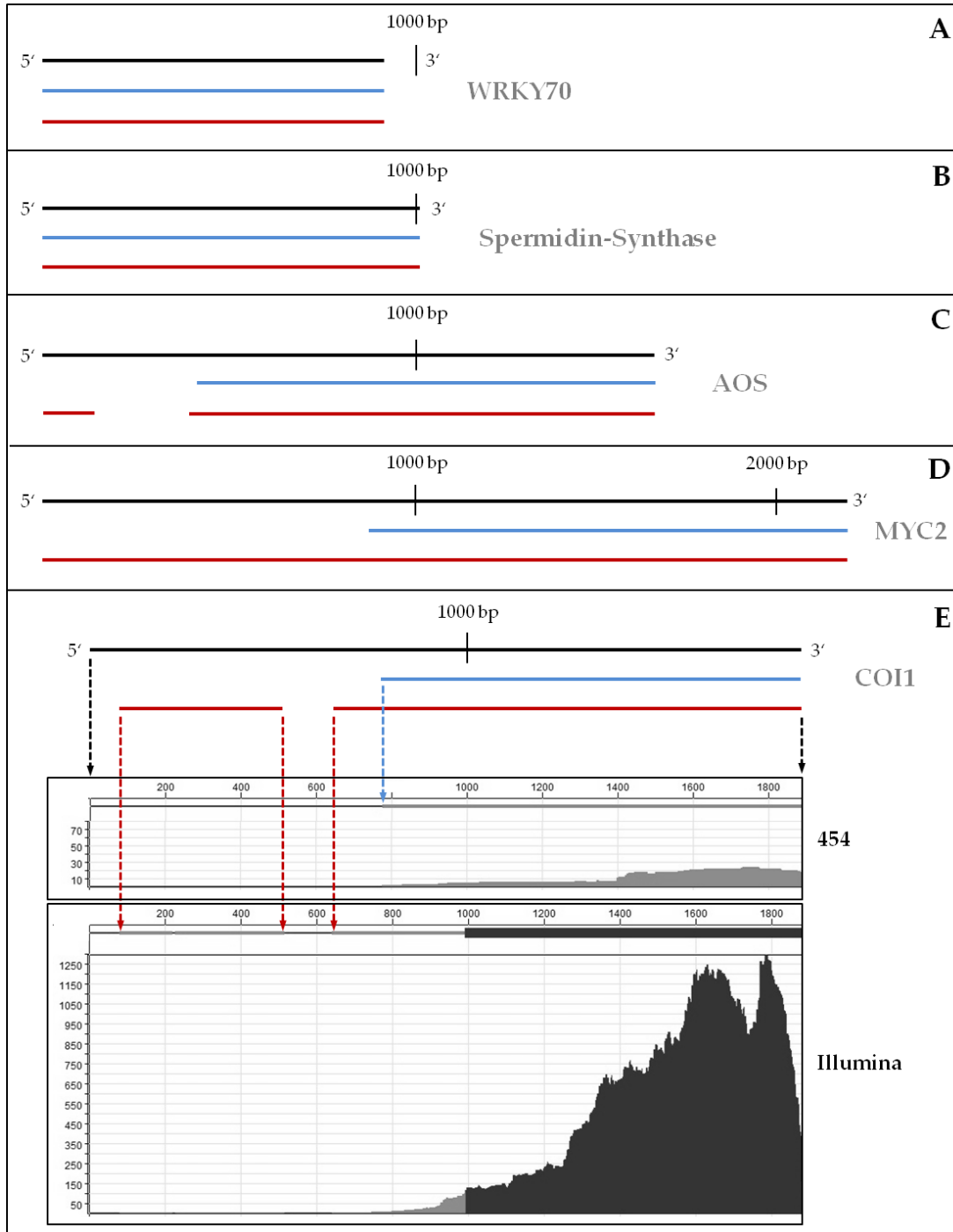


Abbildung 3.3: Abdeckung von ausgewählten Gensequenzen. Dargestellt ist die exemplarische Abdeckung von fünf mRNA-Sequenzen (schwarze Linien) aus der NCBI Datenbank durch 454- (blau) bzw. Illumina-Rohdaten (rot). Die zu Grunde liegende Kartierung der Sequenzdaten wurde mit PASS realisiert, die Assemblierung mit Hilfe der jeweiligen Referenzsequenz mit SeqMan (gezeigt in E, unten). **A** Transkriptionsfaktor WRKY 70 (912 bp) **B** Spermidin-Synthase (1005 bp) **C** AOS – Allene Oxide Synthase (1653 bp) **D** Transkriptionsfaktor MYC2 (2280 bp) **E** COI1 – COronatine Insensitive protein 1 (1760 bp). Auffällig ist die sehr hohe Abdeckung im 3'-Bereich, während in Richtung des 5'-Endes die Zahl der Sequenzen drastisch abnimmt. Des Weiteren sind gerade längere Gene durch die Illumina-Daten besser abgedeckt als durch die 454-Daten (D).

Tabelle 3.4: Redundanz der Illumina-Daten. Gezeigt ist die ursprüngliche und tatsächliche Sequenzinformation, verursacht durch Überlappung der Sequenzen.

	A14	P28
Sequenzinformation (Basen)	8.028.908.497	7.617.721.563
Überlappung der "Paired-End"-Sequenzen (Basen)	43	47
Redundanz (%)	21,5	23,5
tatsächliche Sequenzinformation	6.302.693.170	5.827.556.995

3.3 ASSEMBLIERUNG

Zur besseren Datenverarbeitung wurden die in 3.2 analysierten Rohdatensätze zunächst mit geeigneter Software assembliert. Dies wurde im Falle der Roche 454-Daten mit NGen, im Falle der Illumina-Daten mit ABySS durchgeführt. Die Parameter der ABySS-Assemblierung mussten dabei mehrfach angepasst werden, um die Resultate mit den Ergebnissen der NGen-Assemblierungen vergleichen zu können. Dies ist bei Assemblierungen ohne geeignete Referenz eine gängige Vorgehensweise [165]. Die wichtigsten Resultate sind in Tabelle 3.5 zusammengefasst.

Tabelle 3.5: Vergleich der *de novo* Assemblierungen. Gezeigt sind die Ergebnisse der mit NGen assemblierten Roche 454-Daten im Vergleich zu den mit ABySS assemblierten Illumina-Daten (jeweils A14 und P28). Unterstrichen ist die Zahl an Basen, die in jeder Assemblierung in Contigs zusammengefasst werden konnte. Unabhängig von der Größe der zu Grunde liegenden Sequenzdatenbank und Art der Assemblierung ergibt sich hierbei ein Wert von rund 18 Mb.

Datenquelle Spezies	Roche 454		Illumina	
	A14	P28	A14	P28
Anzahl Contigs	24.620	24.146	30.643	39.194
Anzahl der in Contigs assemblierten Basen	<u>18.038.940</u>	<u>18.087.985</u>	<u>18.192.061</u>	<u>18.416.635</u>
Ø Contiglänge (Basen)	733	749	594	470

Die Assemblierungen der beiden Roche 454-Datensätze mit dem Programm NGen liefern mit 24.620 (A14) und 24.146 Contigs (P28) sehr vergleichbare Werte. Die mittlere Contiglänge ist mit 733 (A14) und 749 Basen (P28) ebenfalls ähnlich. Dagegen weicht die Zahl der Contigs bei den mit ABySS assemblierten Daten trotz Anpassung der Parameter stärker ab (30.643 bei A14 im Gegensatz zu 39.194 bei P28). Auch die mittleren Contiglängen zeigen mit 594 (A14) bzw. 470 Basen (P28) ähnliche Unterschiede. Auffällig ist das umgekehrt proportionale Verhältnis von mittlerer Contiglänge zur Gesamtzahl von Contigs. Trägt man diese Werte gegeneinander auf und ermittelt die Regressionsgerade [SI Tabelle_1], ergibt sich R^2 zu 0,984 und damit einer sehr guten Korrelation zwischen diesen beiden Größen.

Ersichtlich wird dies darüber hinaus an der Anzahl der nicht-redundanten Basen, die zu Contigs assembliert wurden. Diese liegt unabhängig von Größe des Datensatzes und Art der Assemblierung bei rund 18 Millionen Basen (Tabelle 3.5). Aufgrund dieser ähnlichen Werte, gerade vor dem Hintergrund der signifikant unterschiedlichen Datenbankgrößen, kann somit davon ausgegangen werden, dass das Transkriptom von beiden Datensätzen vollständig abgedeckt wird. Die 18 Mb an nicht redundanten, transkribierten Sequenzen entsprechen somit 2,4 % der Genomgröße, allerdings muss berücksichtigt werden, dass der tatsächliche Wert aufgrund der hier beobachteten 3'-Verschiebung (siehe 3.2) vermutlich höher anzusiedeln ist.

Die Contiglängenverteilung zeigt für die beiden mit NGen assemblierten Roche 454-Datensätze nahezu identische Diagramme (Abbildung 3.4). Die zahlenmäßig häufigsten Contiglängen liegen mit 494 (A14) und 485 Basen (P28) nur geringfügig unter dem Maximum der Länge der Rohdaten (vgl. Abbildung 3.2). Insgesamt sind 81 % der Contigs länger als 500 und rund 16 % länger als 1000 Basen (1,5% > 1500; 0,3 % > 2000), in diesem Größenbereich folgen die Kurven somit einer hyperbolischen Funktion.

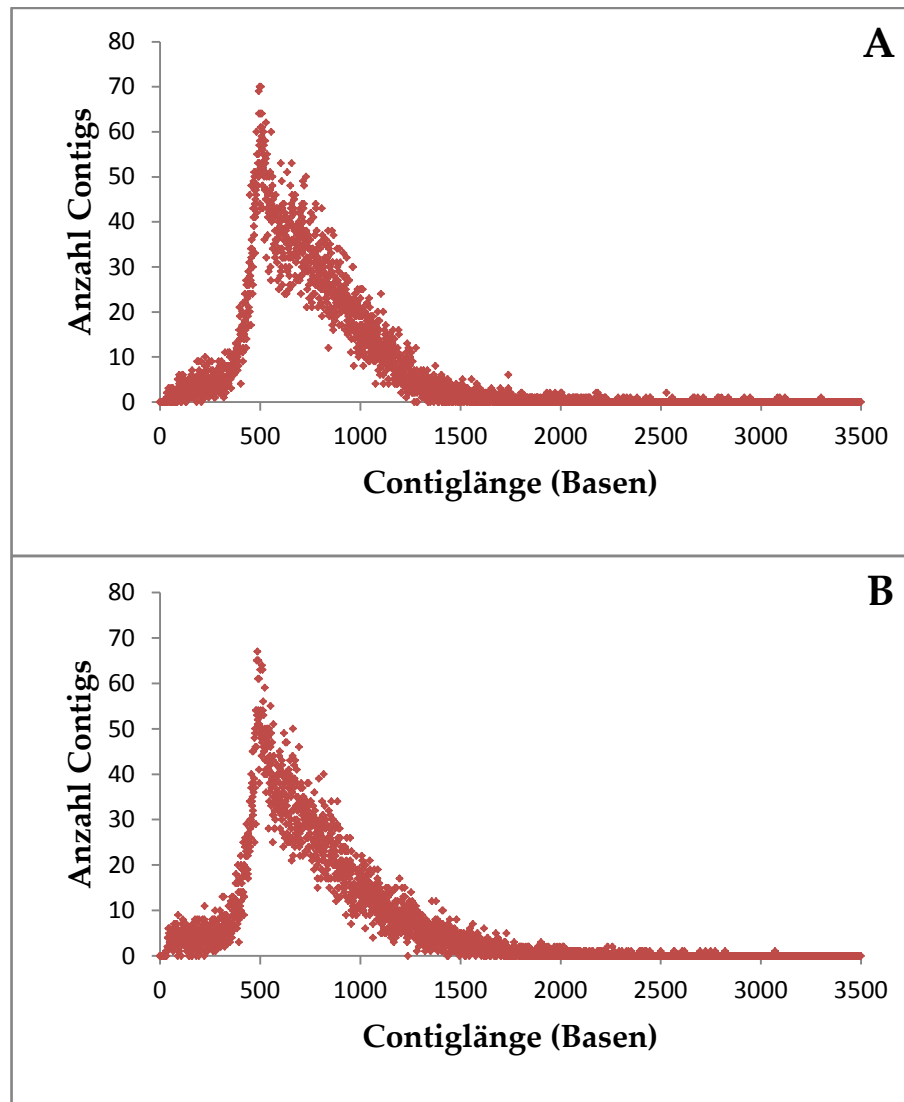


Abbildung 3.4: Contiglängenverteilung nach NGen Assemblierung. Dargestellt ist das Ergebnis der assemblierten 454 Daten von A14 (A) und P28 (B). Die Anzahl der Contigs ist gegen die entsprechende Länge aufgetragen, mittlere Contiglänge ist 732 (A14) und 749 Basen (P28) mit ihren lokalen Maxima bei 494 bzw. 485 Basen.

Die Contiglängenverteilung der mit ABySS assemblierten Illumina-Daten zeigt zwei lokale Maxima bei 83 und 165 Basen (Abbildung 3.5, blaue Pfeile). Dies lässt sich mit dem gewählten Parameter $k = 83$ (Zerlegung der „reads“ in 83-mere) erklären, da dadurch zehn (Coverage $c = 10$) identische bzw. den Assemblierungs-Parametern entsprechend zusammengehörige k -mere bereits einen Contig bilden. Gleichmaßen werden zwei k -mere mit einer Überlappung von einer Base in Contigs der Länge $2k-1 = 165$ Basen zusammengefasst.

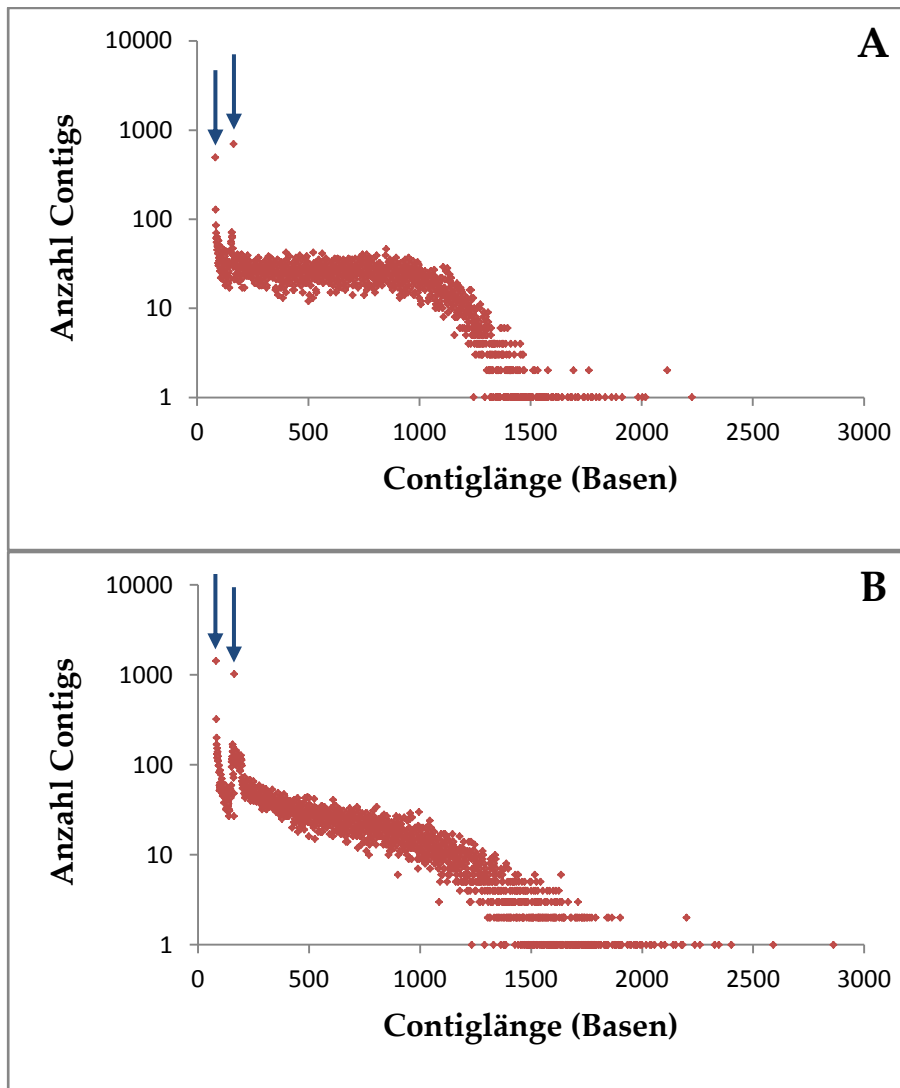


Abbildung 3.5: Contiglängenverteilung nach ABySS Assemblierung. Gezeigt ist das Ergebnis der assemblierten Illumina-Daten von A14 (A) und P28 (B), wobei die Anzahl der Contigs gegen ihre entsprechende Länge aufgeführt ist. Mit zunehmender Länge nimmt die Zahl der Sequenzen in P28 deutlich schneller ab als in A14. Die Pfeile markieren die überrepräsentierten Sequenzen der Länge 83 (= k) und 165 (= 2k-1).

Da dies gegen die gewählten Parameter verstoßen würde, muss hier ebenfalls eine minimale Abdeckung von zehn „reads“ vorliegen. Tatsächlich finden sich bei der eingehenden Analyse der Contigs [SI FASTA_1 und SI FASTA_2] 60 % der Contigs (n = 50) der Länge 83 mit einer geringen Abdeckung von 10-50 k-meren ($\bar{\varnothing}$ 848, n = 50). Dagegen basieren 98 % der Contigs (n = 50) der Länge 2k-1 auf über 1000 k-meren ($\bar{\varnothing}$ 6555, n = 50). Hierdurch ist gewährleistet, dass der Contig auch von k-meren abgedeckt wird, die eine größerer Überlappung als die oben erwähnte eine

Base zeigen, was im Übrigen wiederum den gewählten Parametern widersprechen würde. Der Vergleich der Assemblierungen zwischen A14 (Abbildung 3.5 A) und P28 (Abbildung 3.5 B) liefert in der Gegenüberstellung zu den homogenen Verteilungen der assemblierten Contigs aus den Roche 454-Daten deutlichere Unterschiede. So sind nur 56 % (A14) bzw. 38 % (P28) der Contigs länger als 500 Basen (15 bzw. 10 % > 1000; 0,2 bzw. 1 % > 1500; 0,01 bzw. 0,05 % > 2000). Auffällig ist, dass trotz des niedrigeren Anteils von Contigs über 500 Basen (zurückzuführen auf die höhere Zahl an Contigs) in P28 die fünffache Menge an Contigs über 1500 bzw. 2000 Basen zu finden ist.

Zur weiteren Bearbeitung wurden alle Contigs in FASTA-Dateien [SI FASTA_1 bis SI FASTA_4] zusammengefasst, welche die Länge und Anzahl der assemblierten „reads“ zu jedem Contig beinhalten.

3.4 VERARBEITUNG DER SEQUENZINFORMATIONEN

Dieses Kapitel beschreibt die in dieser Arbeit etablierte und angewandte Form der Verarbeitung transkriptomischer Daten von Nicht-Modellorganismen. Es wurde großer Wert darauf gelegt, eine leicht reproduzierbare Methodik mit weit gefächertem Anwendungsspektrum zu kreieren, was vor allem Datenbank- und Speziesunabhängigkeit beinhaltet. Die Pipeline ist insbesondere darauf ausgelegt, in kurzer Zeit von reinen Sequenzdaten zu biologisch relevanten Informationen zu gelangen. In den folgenden Kapiteln wird nun wiederholt auf die Verwendung unterschiedlicher Datenbanken eingegangen und die hier dargestellten Ergebnisse sind direkt abhängig von Art, Größe und Inhalt der jeweiligen Datenbank. Es soll daher an dieser Stelle darauf hingewiesen werden, dass öffentlich verfügbare Datenbanken einer ständigen Aktualisierung – basierend auf gegenwärtigen Forschungsergebnissen, Umstrukturierungen etc. – unterliegen. Soweit dies möglich ist, werde ich deshalb angeben, von welchem Datum die jeweils verwendete Datenbank stammt.

3.4.1 DATENPIPELINE

Als eine zum Beginn der Arbeit erst aufkommende Technologie stellte die Verarbeitung von Daten der zweiten Generation des Sequenzierens die weitaus

größte Herausforderung dar. Auch heute existieren nur wenige, universell anwendbare Protokolle, weshalb das Prozessieren der Daten auf jede Fragestellung individuell angepasst werden muss. Um das gesteckte Ziel eines umfassenden transkriptomischen Vergleichs mehrerer Datensätze zu realisieren, wurde daher eine eigens dafür vorgesehene Datenpipeline etabliert (Abbildung 3.6) und unter den Gesichtspunkten der Reproduzierbarkeit, des Anwendungsspektrums und der Schnelligkeit sukzessive verbessert.

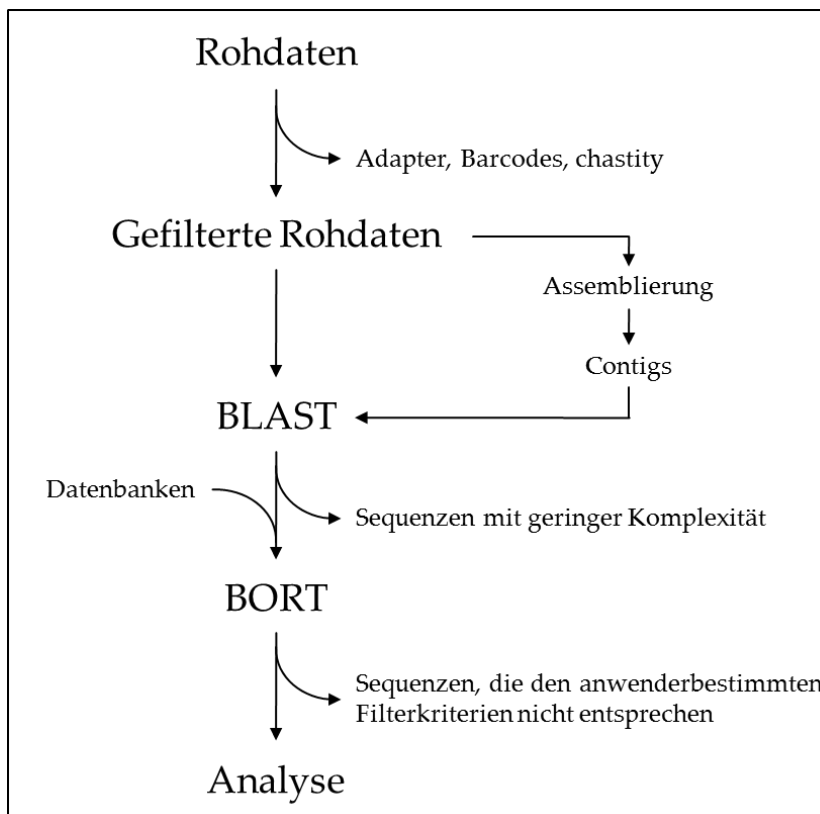


Abbildung 3.6: Datenpipeline. Die Rohdaten werden zunächst von technischen Sequenzen (Vgl. Fußnote 3, Seite 48) befreit. Illumina-Daten werden zusätzlich nach ihrem „chastity“-Wert⁵ gefiltert. Die so erhaltenen Sequenzen können entweder direkt einer BLAST-Suche unterzogen oder vorher zu Contigs assembliert werden. Nach der BLAST-Suche können optional „reads“ bzw. Contigs mit geringer Komplexität herausgefiltert werden, bevor die BLAST-Resultate mit dem zentralen Programm BORT nochmals anwenderspezifisch gefiltert, umformatiert, zusammengefasst und ausgegeben werden.

⁵ Der „chastity“-Wert ist ein Maß für das Abschneiden im Standard Illumina Qualitätsfilter. Dieser Wert (vgl. 2.1) ist definiert als der Quotient aus der höchsten Basenintensität (I_{\max}) und der Summe der höchsten beiden (I_{\max} , $I_{2\max}$). Auf diese Weise können Sequenzen mit niedrigem Signal-Rausch-Verhältnis identifiziert und aussortiert werden.

Zunächst werden die Rohdaten von jeglichen technischen Sequenzen wie Adapter oder Barcodes befreit, Illumina-Daten werden zusätzlich nach ihrem „chastity“-Wert gefiltert (zum Entfernen von Sequenzen mit hohem Signal-Rausch-Verhältnis). Die so gefilterten Rohdaten können dann entweder zu Contigs assembliert werden (vgl. 3.3) oder unbearbeitet für BLAST-Suchen oder Kartierungen eingesetzt werden. Entscheidend ist hierbei die Wahl der Datenbank bzw. einer geeigneten Referenz (Genom). Da dieses zu Beginn der Arbeit noch nicht verfügbar war, wurde anfangs auf Datenbanken wie Reference Sequence (RefSeq) [166] oder die Universal Protein knowledgebase (UniProt) [150, 167] zurückgegriffen. Des Weiteren stand die (Stand März 2010) auf knapp 288.000 ESTs basierende NCBI *Malus* Unigene Datenbank mit rund 23.000 Einträgen zur Verfügung. Ähnlich der RefSeq lassen sich mit dieser zwar insgesamt mehr Treffer (BLAST oder Kartierung) erzielen, jedoch ohne wirklich Informationen zu den betreffenden Genen zu erhalten, da diese nicht ausreichend annotiert sind. Dieses Problem tritt bei Nutzung der UniProt nicht auf, da alle Einträge sehr gut annotiert sind, hier können aber aufgrund der geringen Datenbankgröße weitaus weniger Treffer generiert werden. Mit Veröffentlichung des Genoms von *Malus x domestica* cv. ‚Golden Delicious‘ [9] stand ab September 2010 eine weitere Datenbank mit 57.386 vorhergesagten Genen (*Malus x domestica* Proteinen, MDPs), aber ebenfalls geringem Informationsgehalt zur Verfügung. Für die Analysen zur differentiellen Regulation wurde daher versucht, die Eigenheiten der verschiedenen Datenbanken zu kombinieren, um a) möglichst viele Sequenzen zur statistischen Absicherung nutzen zu können und b) Funktionsanalyse aufgrund zuverlässiger Annotation betreiben zu können. Daher werden in der hier beschriebenen Datenpipeline zunächst alle Sequenzen eines Datensatzes mit den sich als beste Referenz herausgestellten MDPs aus dem Genomprojekt und/oder der NCBI *Malus* Unigene Datenbank abgeglichen. Die BLAST-Resultate werden dann mit dem zentralen Programm der Pipeline, BORT (BLAST Output Refinement Tool) [SI Perl-Skript_6] geparkt. Alle Treffer mit einem E-Wert⁶ > 1E⁻¹⁰ werden dabei herausgefiltert, um unspezifische Alignments zu vermeiden. Weitere Filterkriterien können manuell eingerichtet werden, um individuelle Fragestellungen zu realisieren. Die verbleibenden (signifikanten) Alignments werden anhand ihrer Referenz (description) beim Parsen aufaddiert und man erhält die Zahl der Sequenzen pro Datenbankeintrag („read counts“). Dies

⁶ E-Wert = Erwartungswert bei Benutzung des BLAST-Algorithmus. Dieser Wert gibt an, wie viele gleichgute oder bessere Treffer bei einer beliebigen Abfrage erwartet werden können (je kleiner desto besser).

kann ebenfalls mit Contigs vorgenommen werden, dabei werden die Zahlen der assemblierten „reads“ (im Name des Contigs enthalten) aufaddiert.

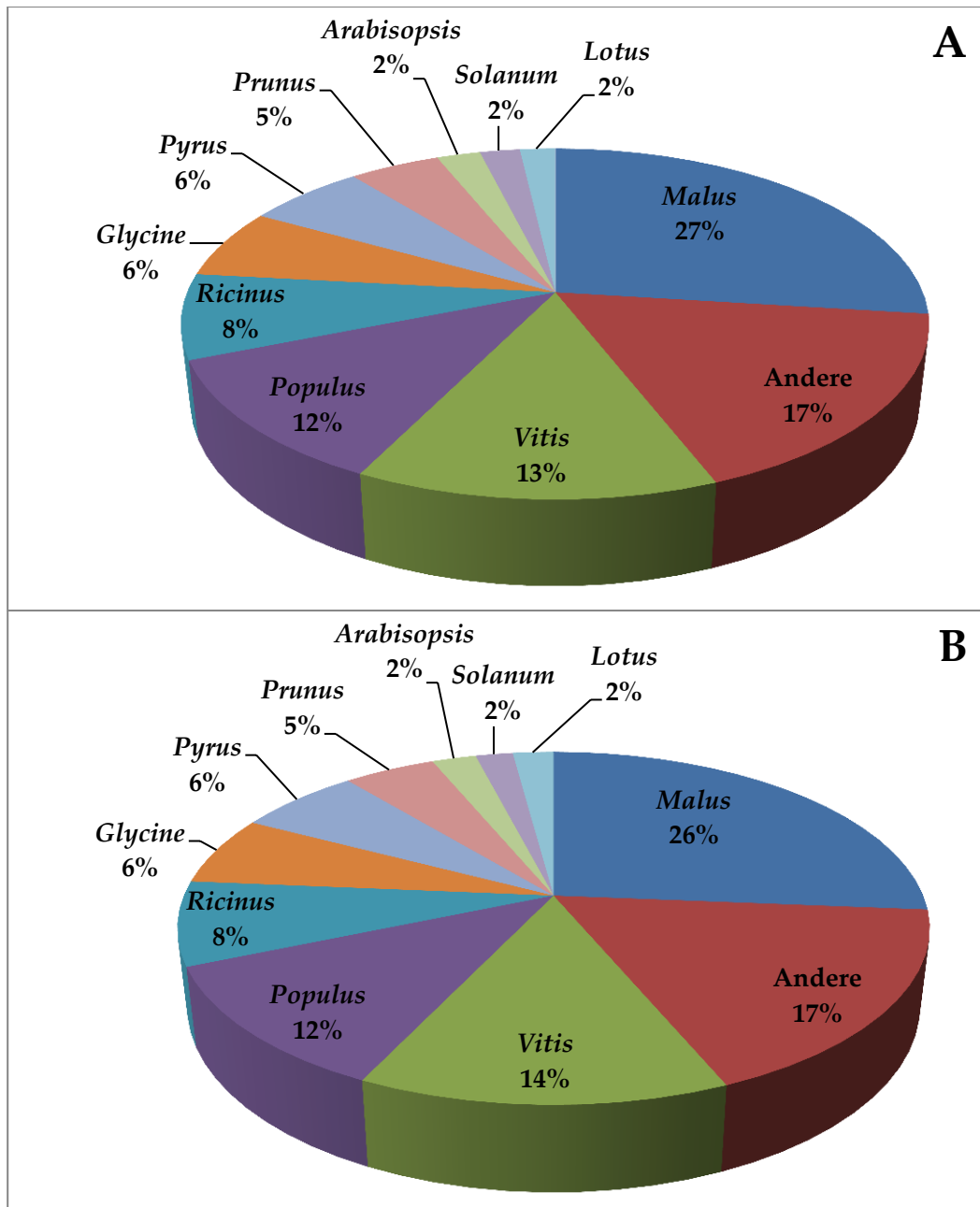


Abbildung 3.7: Speziesverteilung der BLAST-Treffer. Gezeigt ist die prozentuale Verteilung der Datenbanktreffer nach Spezies (je 500.000 BLAST-Treffer aus A14 (A) und P28 (B)). Zu Grunde liegt eine BLASTn-Suche gegen die Nukleotiddatenbank (nt, 03.06.2011) mit einem E-Wert-Grenzwert $\leq 1E^{-10}$. Die beiden Statistiken unterscheiden sich nur minimal, am stärksten vertreten ist erwartungsgemäß die Gattung *Malus* (Ø 26,5 %), gefolgt von *Vitis* (Ø 13,5 %), *Populus* (12 %) und *Ricinus* (8 %). Insgesamt machen die zehn am stärksten repräsentierten Gattungen 83 % aller Treffer aus. Alle Gattungen mit einem Anteil von unter 1 % sind unter ‚Andere‘ (17 %) zusammengefasst.

Diese „read counts“ können vor der weiteren Analyse zusätzlich RPKM-normalisiert⁷ werden, sofern die Länge des Transkripts bekannt ist. Datenbank-einträge mit zugehörigen „read counts“ können dann anwenderspezifisch exportiert werden (TXT oder CSV), wodurch die Größe der zu verarbeitenden Dateien von mehreren Gigabyte (Rohdaten bzw. BLAST-Resultate) auf wenige Megabyte reduziert wird. Die eigentliche Auswertung erfolgt durch den Vergleich mehrerer Datensätze (siehe 3.4.2), die auf die gleiche Art und Weise mit der Pipeline bearbeitet wurden. Die Funktionalität der Pipeline sowie die Art und Verteilung der Organismen, deren annotierte Sequenzen BLAST-Treffer darstellen, wurden durch eine BLASTn-Suche der Illumina-Daten gegen die NCBI Nukleotid (nt) Datenbank (Stand 03.06.2011) durchgeführt, auch hier mit einem E-Wert-Grenzwert von $1E^{-10}$. Das Ergebnis ist für 500.000 BLAST-Treffer in Abbildung 3.7 dargestellt.

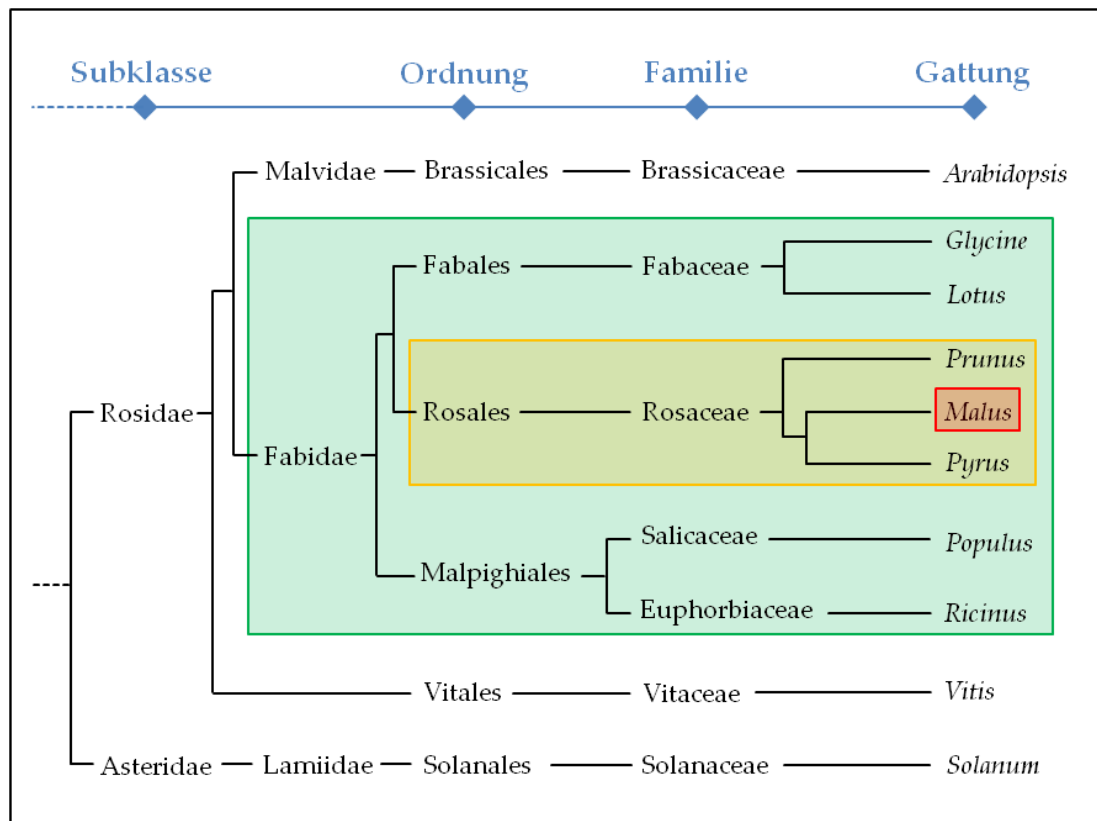


Abbildung 3.8: Kladogramm. Die Grafik zeigt das vereinfachte Kladogramm der zehn Gattungen, die bei BLAST-Suchen von Apfeltranskriptomdaten am häufigsten vertreten sind. Neun von zehn lassen sich in die Subklasse der Rosenähnlichen (Rosidae) gruppieren, davon wiederum sieben zu den Eurosiden I (Fabidae, grün unterlegt). Die Rosenartigen (Rosales) und insbesondere die im Interesse stehende Gattung *Malus* sind hervorgehoben (gelb bzw. rot unterlegt).

⁷ Siehe 2.4.6.1 (Material und Methoden)

Wie erwartet weichen die Ergebnisse der beiden Datensätze nur minimal voneinander ab. Die zehn am stärksten vertretenen Gattungen (Arten) sind demnach *Malus (domestica)* – 26,5 %, *Vitis (vinifera)* – 13,5 %, *Populus (trichocarpa)* – 12 %, *Ricinus (communis)* – 8 %, *Glycine (max)* – 6 %, *Pyrus (pyrifolia)* – 6 %, *Prunus (persica)* – 5 %, *Arabidopsis (thaliana)*, *Solanum (lycopersicum)*, *Lotus (japonicus)* – je 2 %. Die meisten Spezies zeigen einen starken Verwandtschaftsgrad, wie das unter Abbildung 3.8 dargestellte Kladogramm zeigt. Sieben von zehn lassen sich den Eurosiden I (Fabidae) in der Subklasse der Rosidae zuordnen. Lediglich *Solanum* bildet durch die Zugehörigkeit zur Subklasse der Asteridae die einzige Außengruppe.

3.4.2 DIFFERENTIELLE REGULATION

Der zentrale Fokus dieser Arbeit liegt auf der differentiellen Regulation von Genen zwischen SAM der nicht-kolumnaren Sorte A14 und der kolumnaren Sorte P28. In diesem Kapitel soll daher die in dieser Arbeit meistgenutzte Methodik zur Bestimmung differentiell regulierter Gene zwischen zwei Datensätzen anhand der bereits beschriebenen Daten vorgestellt werden. Auf weitere Methoden und die Zusammenführung aller Ergebnisse aus diesen Analysen wird später detailliert eingegangen.

Als Methode der Wahl stellte sich die von Stekel et al. [151] publizierte Formel zum Vergleich der Genexpression in multiplen cDNA-Bibliotheken heraus. Die mit BORT (3.4.1) auf ein einheitliches Format gebrachten BLAST-Ergebnisse wurden dazu in EXCEL importiert und die R_j -Werte als Maß für die Signifikanz der differentiellen Expression berechnet. Anschließend wurden diese Listen nach absteigenden R_j -Werten sortiert und die am stärksten differentiell regulierten Gene eingehend analysiert. Ein Auszug aus einer dieser Listen zeigt Tabelle 3.6, die vollständigen Tabellen für die auf diese Weise bearbeiteten Contigs (Roche 454 und Illumina, vgl. 3.3) und gleichermaßen der Rohdaten (3.2) sind der Arbeit elektronisch angehängt [SI Tabelle_3 bis SI Tabelle_6]. Benutzt man das vorgeschlagene Signifikanzmaß $R_j \geq 8$, ergeben sich für die assemblierten Roche 454-Daten 411, die Rohdaten 148, die assemblierten Illumina-Daten 6920 und für die Rohdaten 4978 differentiell regulierte Gene. Dabei muss allerdings beachtet werden, dass die Berechnung im Falle der assemblierten Illumina-Daten auf k-meren beruht, während sie in den anderen drei Fällen auf „read counts“ basiert.

Tabelle 3.6: Differentiell regulierte Gene. Die unten stehende Tabelle zeigt den Anfang einer insgesamt 9031 Gene umfassenden Tabelle [SI], sortiert nach R_j -Werten als Maß für die Signifikanz der differentiellen Expression. Datengrundlage bilden die mit NGen assemblierten Roche 454-Daten mit angeschlossener BLASTx-Suche gegen die Swiss-Prot Datenbank (Stand 09.03.2011).

Swiss-Prot ID	Beschreibung	#Reads P28	#Reads A14	R_j
O24058	metallothionein-like protein type 2	1271	3928	245,1
P31168	dehydrin cor47	2440	894	204,0
P37392	tubulin beta-1 chain	580	5	178,8
Q9FK81	uncharacterized protein at5g22580	83	977	170,6
P36181	heat shock cognate protein 80	3	444	115,8
:	:	:	:	:

Um herauszufinden, welche Daten sich besser für die Analyse eignen, wurden die Genlisten basierend auf den Contigs [SI Tabelle_3 und 4] und die Genlisten basierend auf den Rohdaten [SI Tabelle_5 und 6] jeweils untereinander verglichen. Der Vorteil der Assemblierung wäre eine weitaus kleinere und besser handhabbare Datenmenge, wohingegen Rohdaten nicht erst aufwändig assembliert werden müssen, wodurch eine mögliche Fehlerquelle (fehlerhafte Assemblierung) ausgeschlossen werden könnte. Da sowohl die Roche 454- wie auch die Illumina-Daten auf dem gleichen Ausgangsmaterial basieren und daher potentiell die gleichen Ergebnisse zu erwarten sind, wurden die ersten 100 Gene jeder Liste für den Vergleich herangezogen. Abbildung 3.9 zeigt, dass die direkte Verwendung der Rohdaten ohne vorherige Assemblierung weitaus bessere Ergebnisse liefert, da sich die auf diese Art generierten Listen wesentlich ähnlicher sind. So zeigen 64 von 100 Genen Abweichung von weniger als 100 Positionen zwischen den beiden Tabellen, nur drei weichen mehr als 1000 Positionen ab. Die weiterbearbeiteten Contigs hingegen zeigen große Unterschiede zwischen den Roche 454- und Illumina-Daten, ausgedrückt durch 33 von 100 Genen, die mehr als 1000 Positionen abweichen und nur 31 von 100 mit geringer Abweichung (> 100 Positionen). Anhand dieser eindeutigen Ergebnisse wurde entschieden, für die Auswertung der verschiedenen Datensätze ausschließlich Rohdaten mit einem hohen E-Wert-Grenzwert von $\leq 1E^{-10}$ ohne vorherige Assemblierung einzusetzen, um die Vergleichbarkeit der produzierten Ergebnisse zu gewährleisten.

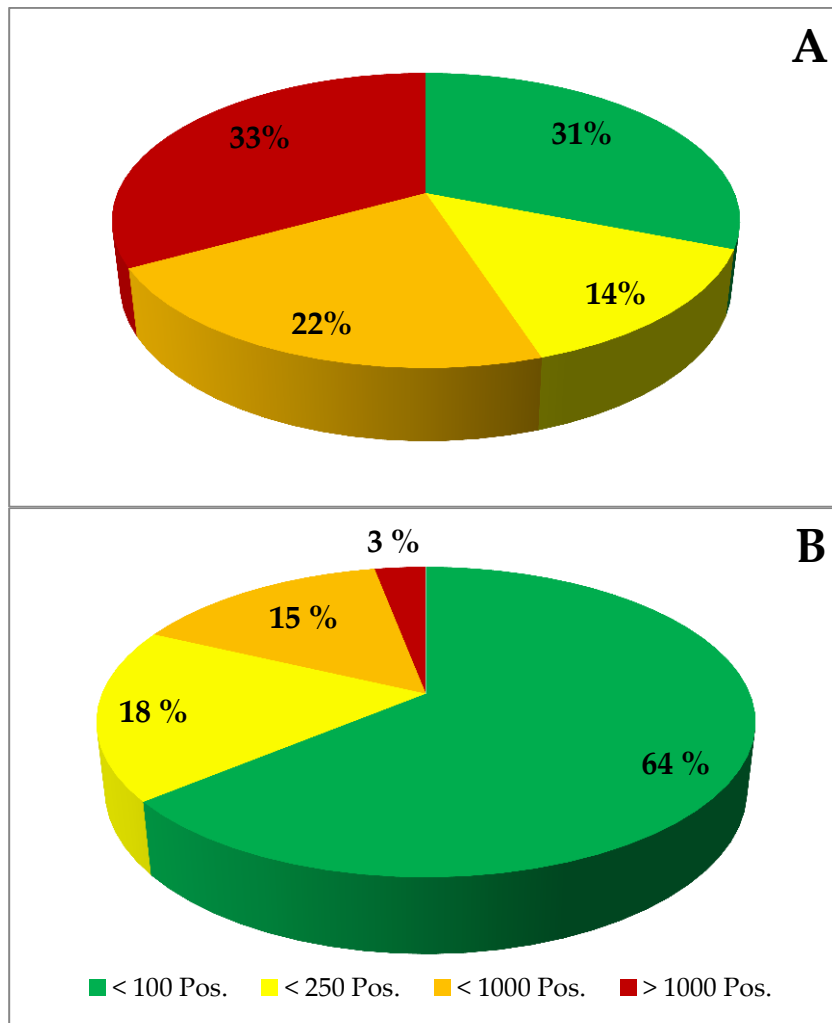


Abbildung 3.9: Contigs versus Rohdaten. Gezeigt ist ein Vergleich der Ergebnisse der mit der vorgestellten Pipeline generierten Genlisten (Top 100) basierend auf den assemblierten Roche 454- und Illumina-Daten (**A**) und der prozessierten Rohdaten ohne vorhergehende Assemblierung (**B**). Die Farbskala gibt dabei den Abstand (in Positionen) der Gene in den beiden gegenüber gestellten Listen an: 0-100 Positionen – grün, 100-250 Positionen – gelb; 250-1000 Positionen – orange, > 1000 Positionen – rot.

3.5 TRANSKRIPTOMDATEN UND KOLUMNARWACHSTUM

Es sollen mit der vorgestellten Methode zur Auswertung der NGS-Daten differentiell regulierte Gene gefunden werden, die unabhängig vom jeweiligen Entnahmezeitpunkt in der kolumnaren Apfelsorte P28 hoch oder herunter reguliert vorliegen und somit potentiell direkt oder indirekt vom Genprodukt des *Co*-Gens in ihrer Expression beeinflusst werden.

3.5.1 GLOBALER VERGLEICH DER DATENSÄTZE

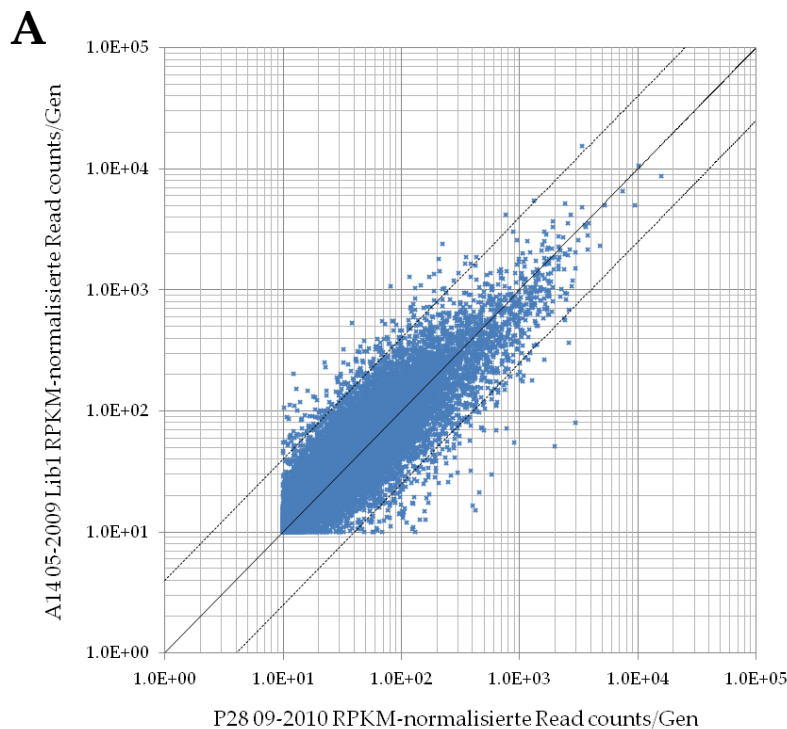
Ein Vergleich multipler Datensätze gestaltet sich in der Regel aufgrund der großen Datenmenge recht schwierig, ist aber sinnvoll, um die Datenqualität zu überprüfen. Eine gute Möglichkeit hierzu bietet die Berechnung der Korrelationskoeffizienten nach Pearson als ein Maß für die Ähnlichkeit zweier Matrizen, in diesem Fall Datensätzen. Dazu wurde eine BLAST-Suche aller Daten gegen die NCBI *Malus* Unigene Datenbank durchgeführt und die „read counts“ wurden für jeden Eintrag RPKM-normalisiert. Auf der Basis dieser Werte wurde die Pearson-Korrelation für alle Datensatzpaare berechnet und gegeneinander gruppiert (Abbildung 3.10).

A14 05-2009 Lib1	A14 05-2009 Lib2	A14 07-2010	P28 05-2009	P28 09-2009	P28 07-2010	
1.00000	0.84646	0.43815	0.85875	0.73407	0.69635	A14 05-2009 Lib1
	1.00000	0.49674	0.84002	0.79978	0.73617	A14 05-2009 Lib2
		1.00000	0.53241	0.35540	0.69382	A14 07-2010
			1.00000	0.72605	0.80468	P28 05-2009
				1.00000	0.72381	P28 09-2009
					1.00000	P28 07-2010

Abbildung 3.10: Korrelationskoeffizienten nach Pearson. Gezeigt ist die Pearson-Korrelation aller Illumina-Sequenzierungen der Transkriptome. Die beiden A14-Bibliotheken des Entnahmezeitpunktes Mai (A14 05-2009 Lib1 und Lib2) entsprechen dabei nicht technischen Replikaten. Die farbliche Unterlegung der Koeffizienten ist gemäß einer Skala von 0 (weiß) bis 1 (blau).

Dabei ergeben sich aufgrund starke Korrelationen (Werte $> 0,5$ zwischen den Transkriptomen der beiden Apfelsorten A14 und P28 und zwischen den verschiedenen Entnahmezeitpunkten (Korrelationskoeffizienten zwischen 0,69 und 0,86). Auffällig ist lediglich der A14-Datensatz des Entnahmezeitpunktes Juli 2010 (**A14 07-2010**), da dieser nur mit dem korrespondierenden Datensatz aus P28 (**P28 07-2010**) eine hohe Ähnlichkeit (0,69), gegenüber allen anderen Datenätzen jedoch deutliche Abweichungen zeigt (\emptyset 0,45).

Ausgewählte Streudiagramme (Abbildung 3.11) zeigen ebenfalls Gemeinsamkeiten sowohl bei Vergleichen (jeweils A14 und P28) von unterschiedlichen (Abbildung 3.11 **A**, Pearson-Korrelation 0,73) als auch gleichen Entnahmezeitpunkten (Abbildung 3.1 **B** und **C**, Pearson-Korrelation 0,84 bzw. 0,69). Dabei wird jedoch auch die große Zahl differentiell regulierter Gene zwischen A14 und P28 deutlich sichtbar. Dies sind in Abbildung 3.11 **A** 368, in **B** (Entnahmezeitpunkt Mai) 407 und in **C** (Entnahmezeitpunkt Juli) 1627, wobei das Streudiagramm bei kleinen RPKM-Werten hier eine klare Verschiebung der Datenpunkte in Richtung A14 zeigt.



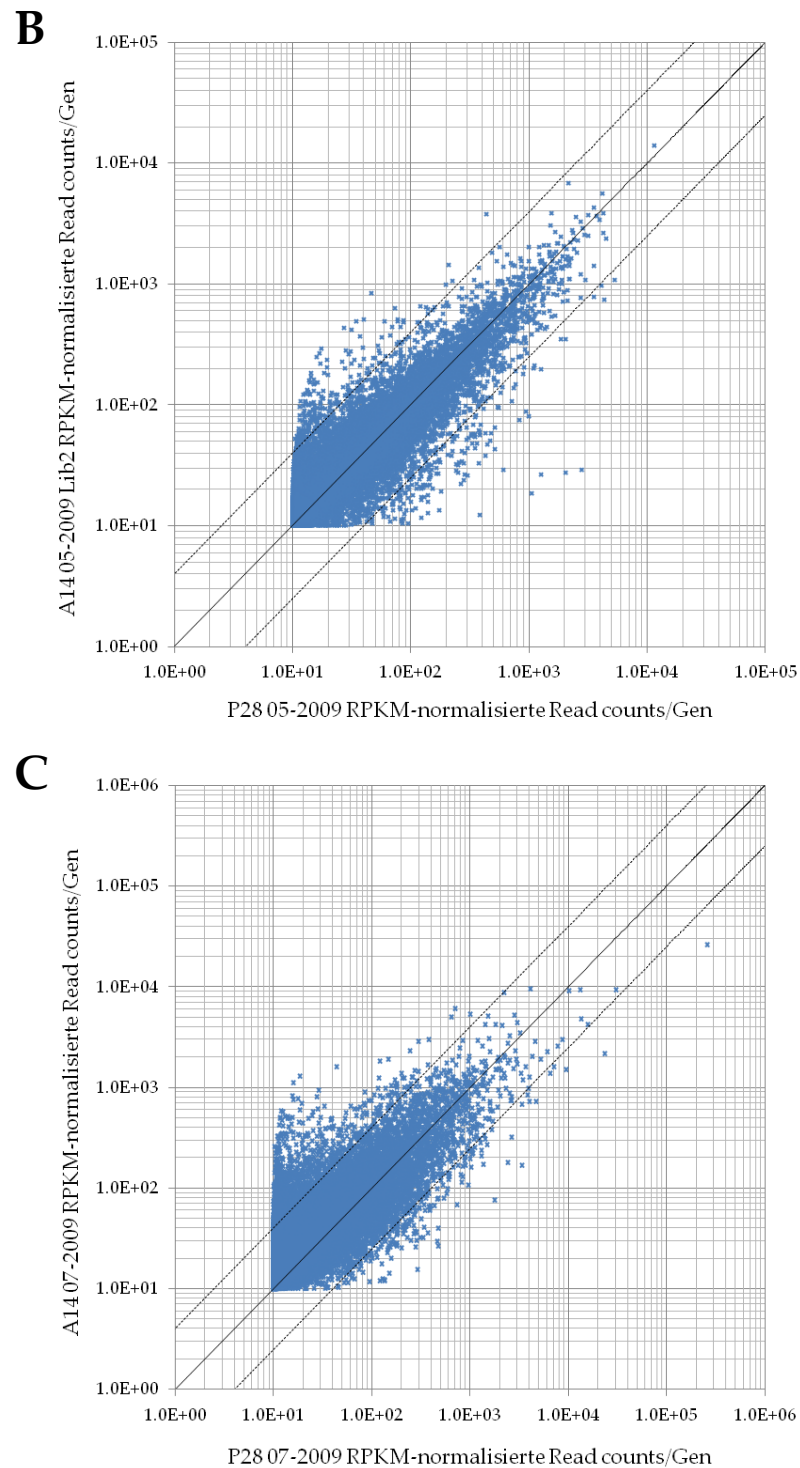


Abbildung 3.11: Streudiagramme RPKM-normalisierter Illumina-Daten. Aufgetragen sind jeweils RPKM-normalisierte „read counts“ von A14 gegen P28, als Referenz diente die NCBI *Malus* Unigene Datenbank. **A** Datensätze vom Entnahmezeitpunkt Mai 2009 (A14) und September 2009 (P28). **B** Datensätze vom Entnahmezeitpunkt Mai 2009 (A14 & P28). **C** Datensätze vom Entnahmezeitpunkt Juli 2009. Zusätzlich zur Null-Linie symbolisieren die beiden weiteren Geraden (gestrichelt) eine vierfach differentielle Regulation.

Da dies noch keinen Aufschluss über die im Fokus stehenden, dauerhaft differentiell regulierten Gene gibt, wurden die in Abbildung 3.11 B und C als stark differentiell reguliert identifizierten Gene in vier Kategorien eingruppiert. So sind in Kategorie I in P28 herunter regulierte, in Kategorie II zum Entnahmezeitpunkt Mai hoch regulierte, in Kategorie III in P28 hoch regulierte und in Kategorie IV zum Entnahmezeitpunkt Juli hoch regulierte Gene. Die Ergebnisse sind in Abbildung 3.12 und detailliert in [SI Tabelle_7] zusammengefasst.

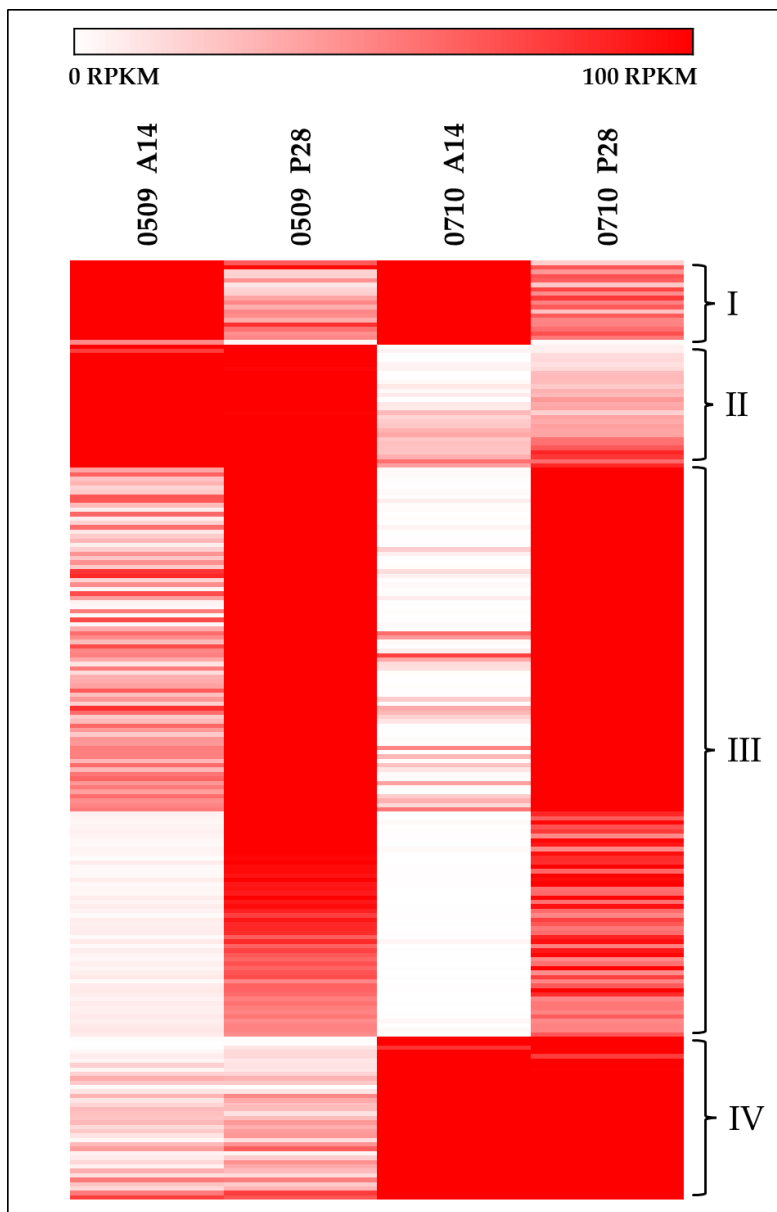


Abbildung 3.12: Genotypisch und entnahmespezifisch differentiell regulierte Gene. Gezeigt ist der paarweise Vergleich der Genotypen A14 und P28 von zwei verschiedenen Entnahmezeitpunkten. Zusammengefasst sind Gene aus vorhergehenden Analysen, die in P28 hoch oder herunter reguliert sind (III bzw. I). Weiterhin sind zwischen den beiden Zeitpunkten Mai bzw. September differentiell regulierte Gene (II bzw. IV) gezeigt. Die farbliche Unterlegung ist gemäß den errechneten RPKM-Werten von 0 (weiß) bis 100 (rot). Die vollständige Tabelle mit Gen-IDs, RPKM-Werten und zugehörigen MapMan-Kategorien ist unter [SI Tabelle_7] abgelegt.

Auffällig ist die hohe Anzahl von 129 Genen, die zu beiden Zeitpunkten signifikant (größer vierfach) in P28 hoch reguliert sind, im Gegensatz zu nur 19 herunter regulierten. Die Anzahl der Gene, die zum Zeitpunkt Mai bzw. Zeitpunkt Juli stärker exprimiert sind, ähneln sich mit Werten von 28 (Mai) bzw. 38 (Juli). Die funktionelle Analyse der Gene mit MapMan [SI Tabelle_7] zeigt, dass sich die in P28 stärker exprimierten Gene vornehmlich den folgenden funktionellen Gruppen (BINs) zuordnen lassen: (1) Zellwandvorstufen und -modifikation, (2) Lipid-Metabolismus, (3) Aminosäure-Metabolismus, (4) Hormonbiosynthese und -signaltransduktion, (5) Transkriptionsregulation, (6) Proteinbiosynthese und -degradation, (7) Zellorganisation und (8) Transport. Die in P28 schwächer exprimierten Gene zeigen aufgrund ihrer geringen Zahl keine Anreicherung in bestimmten funktionellen Gruppen. Auch die Gegenüberstellung differentiell regulierter Gene zwischen Mai und Juli gestaltet sich schwierig, da von den 38 im Juli stärker exprimierten Genen nur 18 (47 %) funktionell charakterisiert werden können. Im Mai sind dies 25 (89 %). Die einzige hier herausstechende Gruppe ist BIN ‚Stress biotisch‘.

Aufgrund der offensichtlichen Unterschiede zwischen den Transkriptomen von A14 und P28, wobei die Mehrheit der differentiell regulierten Gene in P28 höher exprimiert sind, lohnt sich ebenfalls der Blick auf die zu jedem der drei Entnahmezeitpunkte höchst exprimierten Gene (Tabelle 3.7). Diese liegen über die gesamte Wuchsperiode jeweils im Abstand von zwei Monaten auseinander (22.05. – 20.07. – 29.09.). Da diese Zeitpunkte unterschiedliche Wuchsbedingungen beinhalten (Tagesdauer, Temperatur, Niederschlag u.a.), kann hier von einer wachstumsphasenabhängig unterschiedlichen Genexpression ausgegangen werden. Leider war hier der direkte Vergleich mit A14 nicht möglich, da A14-Proben nur für Entnahmezeitpunkt Mai und Juli sequenziert wurden. Bei der Analyse der P28-Transkriptome zeigt sich, dass unter den 10 am stärksten exprimierten Genen jedes Zeitpunktes fünf (S51758971, S26344821, S34535900, S51770972, S26339415) dauerhaft hoch exprimiert sind. Zwei Gene (S21625145, S20630041) sind im Mai und Juli stärker exprimiert als im September, zwei (S51788421, S20636073) im Juli und September stärker als im Mai, drei (S51791243, S51808936, S51789417) nur im Juli und eines (S21624096) ist nur im September stark exprimiert.

Tabelle 3.7: Liste der 10 meistvertretenen Transkripte. Gezeigt sind die zehn am stärksten vertretenen, Protein kodierenden Transkripte aller drei Illumina-Datensätze von P28. Diese sind als RPKM-normalisierte Expressionslevel (Top 10 jeweils fett markiert) mit ihrem dazugehörigen Rang (in Klammern) angegeben. Die Funktion entspricht dem signifikantesten Resultat einer BLASTx-Suche gegen die nr-Datenbank (Stand 28.09.2011).

Gen ID [†]	Procats28 Bibliothek			Funktion
	22.05.2009	20.07.2010	29.09.2009	
Mdo#S51758971	14084 (1)	26090 (1)	8771 (3)	type 2 metallothionein
Mdo#S21625145	6798 (2)	9023 (5)	677	metallothionein-like protein
Mdo#S24849483	5556 (3)	2019	4807 (9)	chalcone synthase
Mdo#S26344821	4258 (4)	9498 (2)	10664 (2)	lipid transfer protein
Mdo#S34535900	3847 (5)	9339 (3)	5527 (5)	peroxidase
Mdo#S51808855	3807 (6)	1379	2352	Nucleotide pyrophosphatase
Mdo#S20630041	3783 (7)	5982 (7)	555	Uncharacterized Protein
Mdo#S51770972	3717 (8)	4178	5043 (7)	cytosolic GAPDH
Mdo#S32258275	3667 (9)	2819	2171	dehydrin 2
Mdo#S51758815	3587 (10)	1796	1887	Stem 28 kDa glycoprotein
Mdo#S51788421	2381	9260 (4)	15508 (1)	Uncharacterized Protein
Mdo#S51791243	490	8727 (6)	847	Uncharacterized Protein
Mdo#S51808936	485	5329 (8)	1435	Uncharacterized Protein
Mdo#S20636073	1191	5216 (9)	5217 (6)	Uncharacterized Protein
Mdo#S51789417	1061	5106 (10)	1650	lactoylglutathione lyase
Mdo#S51754328	2635	3241	6619 (4)	lipid transfer protein
Mdo#S26339415	3253	4104	5009 (8)	cyclophilin
Mdo#S21624096	87	262	4200 (10)	hypothetical protein

[†] Gekürzte Gen ID aus der *Malus x domestica* (Mdo) Unigene Datenbank (Stand 27.06.2011)

3.5.2 KOMPARATIVE GENEXPRESSIONSSTUDIEN

Um alle Datensätze gleichermaßen mit in die Bestimmung differentiell regulierter Gene zwischen A14 und P28 – unabhängig vom jeweiligen Entnahmezeitpunkt des zu Grunde liegenden Materials – mit einbeziehen zu können, wurden alle sequenzierten Bibliotheken mit DESeq [160] analysiert. Dazu wurden die „read count“-normalisierten Illumina-Datensätze von A14 als Referenz, die von P28 als

Probe definiert. Die Analyse aller in dieser Arbeit generierten Datensätze liefert 1173 Gene, die mit einer Fehlerwahrscheinlichkeit $p < 0,005$ und einem „fold change“ $> 3,0$ differentiell reguliert sind [SI Tabelle_8]. Aufgrund dieser für detaillierte Analysen immer noch hohen Anzahl von differentiell regulierten Genen wurden solche mit sehr geringen „read counts“ entsprechend des Grenzwertes $\text{baseMean} < 150$ (entspricht in etwa einem RPKM-Wert von 10) für Folgeanwendungen ausgeschlossen. Nach dem Entfernen dieser weitgehend basal exprimierten Gene, bei denen die Fehleranfälligkeit in Bezug auf differentielle Regulation aufgrund der geringen Anzahl von Sequenzen zudem stark erhöht ist, reduziert sich die Anzahl auf insgesamt 619 Gene. Davon konnten wiederum 465 mit MapMan funktionell charakterisiert werden, die Ergebnisse sind in Abbildung 3.13 und [SI Tabelle_9] dargestellt.

Bei der Analyse ergeben sich sieben überrepräsentierte funktionelle Gruppen (BINs), die nach Wilcoxon-Rangsummentest [168, 169] signifikant sind ($p < 0,05$). Dies sind (1) Stress (biotisch), (2) Proteinmodifikation, -biosynthese und -degradation, (3) Zellwanddegradation (4), Aminosäuremetabolismus (Asparagin), (5) Regulation der Transkription, (6) mitochondrialer Elektronentransport und (7) Transport. Der Großteil der Gene, assoziiert mit vier BINs (1, 3, 4, 6), sind in P28 herunter reguliert, die übrigen drei BINs (2, 5, 7) mit der weitaus höheren Anzahl an Genen sind überwiegend in P28 hoch reguliert. Die beiden BINs ‚Hormone‘ und ‚Regulation‘ wurden trotz fehlender statistischer Signifikanz aufgrund ihrer direkten Assozierbarkeit mit Wachstumsprozessen eingehender analysiert.

Auf die Regulation von Genen, die direkt mit dem Hormonhaushalt der Zelle, also Biosynthese, Signaltransduktion und Transport assoziiert werden können, soll an späterer Stelle eingegangen werden, da hierbei die Erkenntnisse aus Microarrays und qRT-PCR ebenfalls eine wichtige Rolle spielen. Um mögliche Einflüsse signifikant differentiell regulierter Transkriptionsfaktoren (TFs) oder regulationsassoziiierter Gene (BIN ‚Regulation‘) sinnvoll bewerten zu können, wurde zunächst eine Übersicht aller im SAM exprimierten TFs (Abbildung 3.14 A) erstellt. Hierzu wurden alle, auch basal exprimierten Gene ($\text{RPKM} > 0$) aus den sechs Illumina-Datensätzen mit MapMan kartiert und die Anzahl der gefundenen Gene zu den jeweiligen TF-Familien für A14 und P28 gegeneinander aufgetragen. Erwartungsgemäß sind die Werte für beide Sorten bis auf geringe Abweichungen identisch, durch die Präsenz des *Co*-Gens ist somit keine offensichtliche Änderung in der Quantität der exprimierten TFs auszumachen.

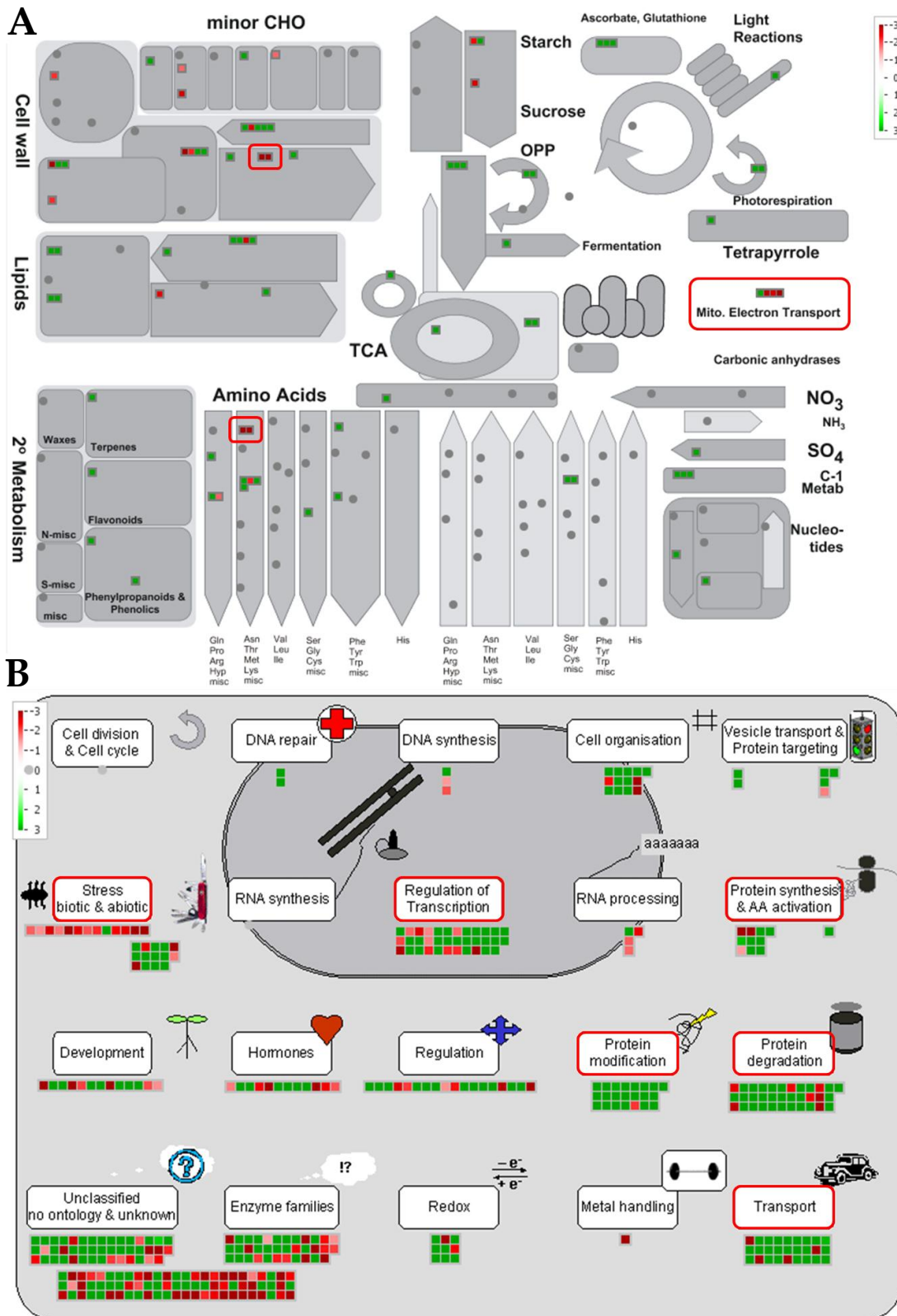


Abbildung 3.13: MapMan-Visualisierung der DESeq Analyse. Gezeigt sind 83 (A) bzw. 391 (B) der funktional gruppierten 465 Gene in ihren jeweiligen BINs. Grün sind in P28 hoch, rot in P28 herunter regulierte Gene. BINs mit $p < 0,05$ sind rot umrandet. Die Skalen geben den Wert der differentiellen Regulation als Log_2 „change fold“ an.

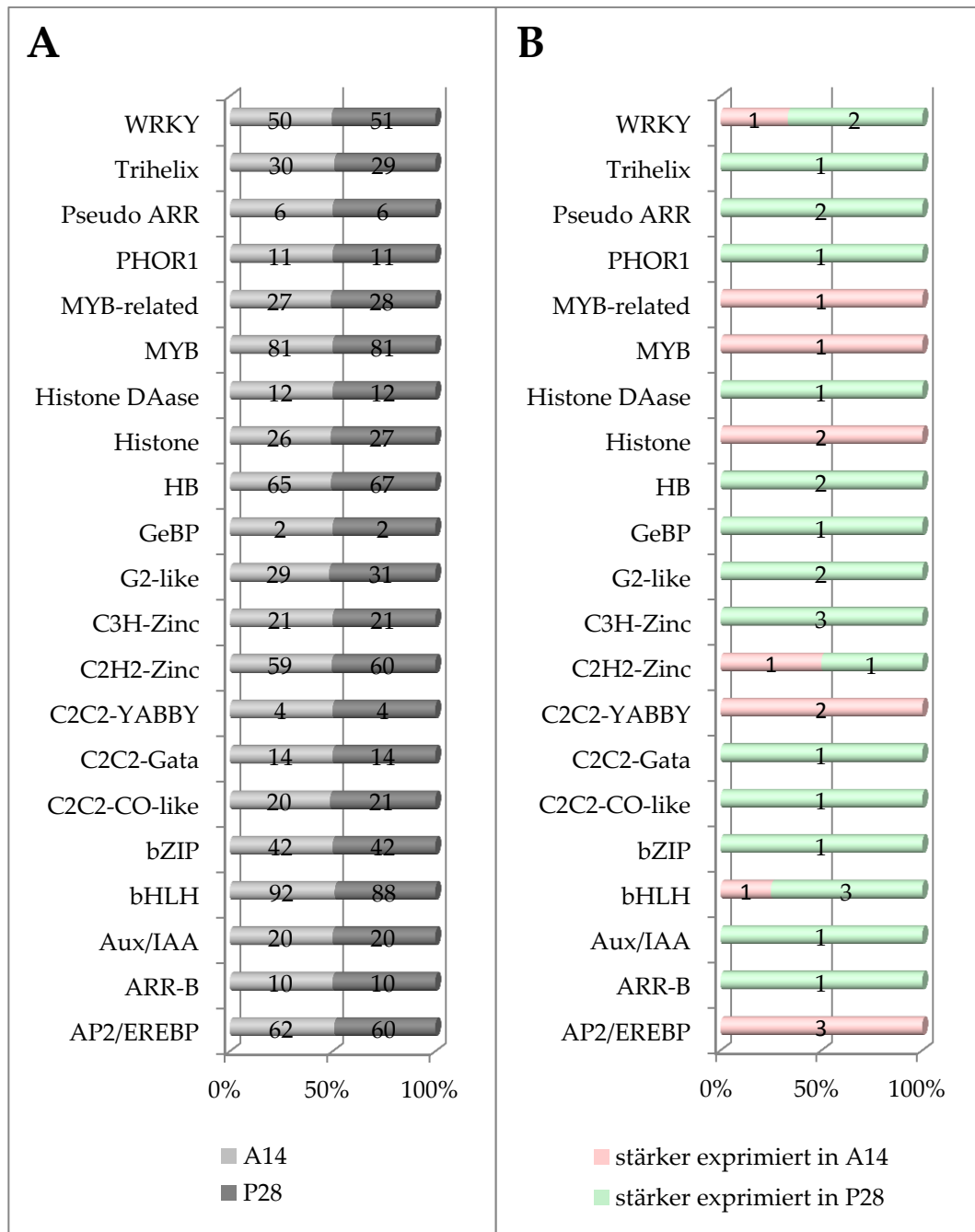


Abbildung 3.14: Differenziell regulierte, TF-assozierte Gene. Gezeigt ist die Anzahl von Genen für die aufgeführten Gruppen von Transkriptionsfaktoren (TFs) oder anderweitig regulationsassoziierten Kategorien, die mit MapMan kartiert werden konnten. **A** Alle im SAM exprimierte Gene aus der Analyse aller sechs Illumina-Datensätze. **B** davon signifikant differenziell regulierte Gene aus der DESeq Analyse.

In den 619 signifikant differenziell regulierten Genen aus der DESeq-Analyse finden sich 36, die sich den aufgeführten Kategorien zuordnen lassen (Abbildung 3.14 B). Davon sind 12 herunter, 24 in P28 hoch reguliert. Besonders interessant erscheinen

Familien, bei denen das Verhältnis von signifikant regulierten zu generell exprimierten hoch ist, wie das bei der Pseudo ARR-Familie (33 %) [170], der GeBP-Familie (50 %) [171] und der C2C2-YABBY-Familie (50 %) [172] der Fall ist, welche alle eng mit Wachstum und Entwicklung assoziiert sind. Weitere interessante Genfamilien stellen AP2/EREBP [173] und ARR-B dar, deren Mitglieder sich mit dem Hormonhaushalt in Verbindung bringen lassen, sowie die PHOR1-Familie, deren TFs in proteasomaler Degradation involviert sind.

Um die bisherigen Erkenntnisse aus der Analyse der NGS-Daten zu verifizieren und drei weitere Entnahmezeitpunkte in Betracht zu ziehen, wurden zusätzliche Expressionsstudien in Form von Microarrays und qRT-PCRs vorgenommen.

3.5.2.1 *Microarrays*

Die Auswahl der Gene für die Microarrays [SI Tabelle_10] erfolgte leider zu einem Zeitpunkt, zu dem die NGS-Daten noch nicht vollständig in verwertbarer Form vorlagen. So deckt die Genauswahl als Sonden auf dem Microarray die per NGS identifizierten, differentiell regulierten Gene [SI Tabelle_8] nicht perfekt ab.

Bei der Auswertung dieser Microarrays im Hinblick auf differentiell regulierte Gene zwischen A14 und P28 zu den vier verschiedenen Entnahmezeitpunkten (06.04.10 – 20.04.10 – 04.05.10 – 22.05.09) wird deutlich, dass die Regulation vieler Gene wie auch bei den NGS Daten abhängig vom jeweiligen Entwicklungszustand des Baumes ist [SI Tabelle_11], sich also nicht für jeden beobachteten Entnahmezeitpunkt reproduzieren lässt. Es wurden daher im Folgenden nur Gene betrachtet, die zu allen gemessenen Zeitpunkten und für mindestens eine der beiden generierten Sonden eine – analog zur DESeq-Analyse der NGS-Daten – einheitliche Regulation zeigen. Dies sind insgesamt 55 in P28 hoch regulierte und 36 in P28 herunter regulierte Gene [Grüne und rote Umrahmung in Tabellenblatt „Sonden_extr“, SI Tabelle_11].

Tabelle 3.8: Microarray-Auswertung. Die Tabelle beinhaltet einen Auszug aus den Microarray-Ergebnissen. Gezeigt sind Gene, die zu allen gemessenen Zeitpunkten (1* = 6.4.2010; 2* = 20.4.2010; 3* = 4.5.2010; 4* = 22.5.2009) eine signifikante (change fold ≥ 2) Regulation in mindestens einer der beiden Sonden zeigen. In P28 im Vergleich zu A14 hoch regulierte Gene sind grün, herunter regulierte rot, nicht signifikant regulierte weiß gefärbt. Als Vergleich dazu ist die differentielle Regulation in den verschiedenen NGS-

Datensätzen (1 = 22.5.2009 (A14, P28) – 454, 2 = 22.5.2009 (A14) und 29.7.2009 (P28) – 454, 3 = 22.5.2009 (A14) und 29.7.2009 (P28) – Illumina, 4 = 22.5.2009 (A14, P28) – Illumina, 5 = 20.7.2010 (A14, P28) – Illumina) mit gleicher Farbkodierung gezeigt. Weiterhin sind die Ergebnisse von BLASTn-Suchen der Illumina-Daten (4) ausschließlich gegen die Sondensequenzen (je 50 nt;†) und die Ergebnisse relativer qRT-PCRs (n = 3;‡) als Log₂ „change fold“ dargestellt. MDP – *Malus x domestica* Protein; NGS – Next Generation Sequencing; n.b. – nicht bestimmt.

MDP	Sonde	Beschreibung	Microarray				†	NGS					‡
			1*	2*	3*	4*		1	2	3	4	5	
MDP0000679810	F-650	-				5,9						6,0	
	F-700												7,3
MDP0000909874	F-1350	Cytokinin hydroxylase CYP35A1				4,8						2,8	
	F-1500												4,2
MDP0000146076	F-1200	Probable trans-2-enoyl-CoA reductase, mitochondrial				0,8						0,4	
	F-1250												0,0
MDP0000565104	F-1750	Probable peptide/nitrate transporter At5g13400				1,1						1,2	
	F-1800												1,7
MDP0000613462	F-1050	interferon-induced GTP-binding protein mx, putative				1,3						2,3	
	F-1100												2,2
MDP0000276057	F-1700	UDP-glycosyltransferase 87A1				4,0						1,7	
	F-1750												3,0
MDP0000247363	F-400	-				5,8						4,7	
	F-450												6,2
MDP0000871409	F-600	Protein TIFY 10A				0,4						-0,6	
	F-700												0,6
MDP0000834156	F-450	Aldehyde dehydrogenase family 2 member B7, mitochondrial				-0,7						0,0	
	F-500												0,7
MDP0000504527	F-1350	Epidermis-specific secreted glycoprotein EP1				-2,2						-1,4	
	F-1450												-0,9
MDP0000309906	F-4300	Kinesin-4				0,6						0,3	
	F-4350												0,0
MDP0000268093	F-1400	High mobility group B protein 13				0,2						1,3	
	F-1450												0,2
MDP0000192235	F-1250	Peroxidase 42				1,9						1,5	
	F-1300												2,1
MDP0000129445	F-1150	U-box domain-containing protein 21				-0,3						n.b.	
	F-1250												-1,1
MDP0000084546	F-2200	Cycloartenol synthase				-0,3						n.b.	
	F-2250												-0,9
MDP0000122958	F-250	Thaumatococcus-like protein 1b				-1,5						-1,3	
	F-450												-1,0
MDP0000607000	F-3600	UPF0481 protein At3g47200				-1,8						n.b.	
	F-3650												-0,7
MDP0000670256	F-1800	Peroxisomal acyl-coenzyme A oxidase 1				2,3						n.b.	
	F-1950												2,3

† BLASTn der Illumina-Daten (A14 und P28) gegen die Sondensequenzen (50 nt) - LOG₂ Readcount change fold

‡ Ergebnis der relativen qRT-PCR auf *in vitro*-Kulturen von A14 und P28 - LOG₂ change fold

3.5.2.2 qRT-PCRs

Die qRT-PCRs ($n \geq 3$) wurden aufgrund der Unterschiede in den Expressionsniveaus verschiedener Entnahmezeitpunkte (vgl. NGS und Microarrays) auf der Basis von *in vitro*-Kulturen von A14 und P28 vergleichend durchgeführt [174], da diese konstanten Kulturbedingungen unterliegen und somit der Entnahmezeitpunkt keine übergeordnete Rolle bei der Regulation gemessener Gene spielen sollte.

Die Ergebnisse aller drei Methoden im Vergleich mit Fokus auf die Microarray-Resultate zeigen Tabelle 3.8 und SI Tabelle_11. Hieraus geht hervor, dass die Ergebnisse der verschiedenen Methoden zur Expressionsanalyse teilweise sehr gut übereinstimmen, andererseits aber auch signifikante Unterschiede erkennbar sind. Für acht von dreizehn Genen bestätigen alle drei Methoden eine Hochregulierung in P28, für vier zeigen qRT-PCR und/oder NGS-Daten keine oder entgegengesetzte Regulation, für ein Gen (Epidermis-specific secreted glycoprotein EP1) besteht eine Diskrepanz zwischen NGS-/qRT-PCR-Daten und Microarray-Daten. Auch in laut Microarrays in P28 herunter regulierten Genen gibt es Übereinstimmungen und Diskrepanzen. Es konnten nur Ergebnisse für zwei von fünf Genen mit Hilfe von NGS-Daten und qRT-PCR als übereinstimmend eingeordnet werden (in Tabelle 3.8 ist lediglich eines gezeigt). Aufgrund der Tatsache, dass Illumina-Daten große Teile der kodierenden Sequenz eines Gens abdecken, während die Microarray-Analysen auf zwei Sonden (50 nt) im 3'-UTR oder im letzten Exon basieren, wurden alle Illumina-Daten zusätzlich ausschließlich gegen die Sondensequenzen [SI FASTA_5] abgeglichen. Dies wurde mit stringenten BLAST-Suchen (nur Treffer > 40 nt (80 %) mit Identität > 90 %) realisiert. Dabei wird deutlich († in Tabelle 3.8), dass diese Ergebnisse stark von denjenigen aus der Datenpipeline abweichen können. Auffällig ist jedoch, dass sie fast vollständig mit den Ergebnissen der qRT-PCR übereinstimmen, sowohl in der Tendenz wie auch der Höhe (\log_2 „change fold“) der differentiellen Regulation.

In der Summe zeigen Microarray- und NGS-Analysen gute Übereinstimmungen bei signifikant differentiell regulierten Genen zwischen A14 und P28, unter Einbeziehung von qRT-PCRs lassen sich immer noch 10 von 18 Genen (55 %) als übereinstimmend reguliert bestätigen.

3.5.3 HORMONHAUSHALT IM KOLUMNAREN SPROSSAPIKALMERISTEM

Um die Erkenntnisse aus allen Genexpressionsstudien mit in die Analyse der hormonalen Situation im kolumnaren SAM einfließen zu lassen, soll erst an dieser Stelle näher auf die Phytohormon-Regulation eingegangen werden. Tabelle 3.9 fasst alle signifikant differentiell regulierten Gene zusammen, die sich mit Hormonbiosynthese, -transport und -signaltransduktion assoziieren lassen.

Zusammenfassend lassen sich die Gene in Bezug auf das Kolumnarwachstum in interessante Kategorien einordnen. So sind Lipoxygenase 2 (LOX2), Allene Oxide Synthase (AOS), Allene Oxide Cyclase (AOC4) und Jasmonate Resistant (JAR1) Gene der Jasmonat (JA)-Biosynthese und sämtlich in P28 hoch reguliert. Squalene monooxygenase (SQLE) und Delta(24)-sterol reductase Diminuto/Dwarf 1 (DIM/DWF1) sind wichtige Schaltergene in der Biosynthese von Brassinosteroiden (BRs) und ebenfalls in P28 stärker exprimiert. Touch Protein 4 (TCH4) kodiert für eine Xyloglucan Endotransglucosylase/hydrolase und ist ein BR-abhängiges Effektorgen, in P28 jedoch herunter reguliert. Gibberellin 2-beta-dioxygenase 1 (GA2ox1) reguliert die Synthese bioaktiver, polarer Gibberelline (GA) und ist in P28 weitaus weniger stark exprimiert. Alle mit der Cytokinin-Synthese (Cytokinin hydroxylase – CYP735A1) und –signaltransduktion (Histidine-containing phosphotransfer protein 1 – AHP1; type B phosphoaccepting response regulator – ARR-B) assoziierten Gene sind wiederum in P28 hoch reguliert. Die größte Gruppe stellen die mit dem Phytohormon Auxin (Indol-3-essigsäure, IAA) assoziierten Gene mit sechs Vertretern dar. Dabei sind die am Auxin-Transport beteiligten Gene Auxin efflux carrier component 1 (PIN1) und Auxin transporter protein 1 (AUX1) hoch, Auxin-binding protein 19a (ABP19a) zur Hormonperzeption herunter reguliert. Auxin-responsive protein IAA27 (IAA27) als Repressor früher Auxin-abhängiger Gene ist in P28 stärker, die beiden Mitglieder der Auxin-induzierten GH3-Familie Indole-3-acetic acid-amido synthetase GH3.1 und GH3.2 (GH3.1 und GH3.2) schwächer exprimiert.

Anhand der hier beobachteten Regulation der besagten Gene im SAM kolumnarer Apfelbäume lässt sich demnach auf erhöhte Konzentrationen von JAs (insbesondere das hochaktive JA-Ile durch die Präsenz und Regulation von JAR1), BRs und Cytokinin (vornehmlich trans-Zeatin) schließen. Zudem zeigt die ebenfalls in P28 hoch regulierte UDP-glucosyltransferase 87A1 (UGT87A1) starke Homologien zu einer Zeatin O-Glucosyltransferase aus *Phaseolus lunatus* [175] [SI DOK_1], welche

Tabelle 3.9: Differentiell regulierte, Phytohormon-assoziierte Gene. Die Tabelle umfasst alle signifikant differentiell regulierten Gene aus DESeq- (NGS), Mircoarray- (MA) und qRT-PCR-Analysen (qPCR), die mit Phytohormonhaushalt, -transport und -signaltransduktion assoziiert werden können. Angegeben ist die Gen ID, Bezeichnung (*Populus trichocarpa* Gen ID in Klammern), der Erwartungswert sowie die Regulation als Log₂ „change fold“. In P28 im Vergleich zu A14 signifikant herunter regulierte Gene sind rot, hoch regulierte grün unterlegt. Färbung gemäß der unten angegebenen Farbskala.

Gen ID	Bezeichnung	E-Wert	Herkunft	Regulation
Mdo#S32285858	SQLE	0.0	NGS	3.39
MDP0000753547	LOX2	0.0	qPCR	3.94
MDP0000225501	AOS	0.0	qPCR	0.73
MDP0000276057	UGT87A1	1.0E-177	MA, qPCR	2.80
Mdo#S55982302	GA2ox1	3.0E-171	NGS	-2.50
Mdo#S22218020	ABP19A	1.0E-112	NGS, MA	-1.76
MDP0000469813	AOC4	6.0E-100	qPCR	1.68
Mdo#S26342565	AUX1 (576946)	2.0E-98	NGS, MA	4.85
Mdo#S51814794	CYP735A1	6.0E-95	NGS	4.10
Mdo#S20669658	IAA27	1.0E-90	NGS	4.62
Mdo#S20644799	JAR1 (834775)	2.0E-89	NGS	4.76
Mdo#S20565678	DIM/DWF1	1.0E-82	NGS	5.13
Mdo#S26337816	GH3-2 (821271)	4.0E-69	NGS	-2.70
Mdo#S21701994	GH3-1	9.0E-69	NGS	-2.49
Mdo#S51770377	PYR/PYL (907261)	4.0E-64	NGS	-2.38
Mdo#S51773842	ARR-B (779103)	3.0E-58	NGS	4.57
Mdo#S22328181	TCH4 (570685)	8.0E-55	MA, qPCR	-2.20
MDP0000216664	AHP1	2.0E-52	MA	1.30
Mdo#S32257287	PIN1	6.0E-51	NGS	4.79
Mdo#S20600970	SnRK2 (665248)	4.0E-46	NGS	4.26
Mdo#S32279405	SnRK2 (832255)	1.0E-43	NGS	4.10
Mdo#S20636604	BSK (823961)	2.0E-35	NGS	5.15
Mdo#S20635910	PP2C (768154)	2.0E-32	NGS, qPCR	5.06
Mdo#S20586266	SnRK2 (931343)	4.0E-28	NGS	4.45
Mdo#S32249408	PP2C (1081359)	3.0E-24	NGS	4.32
Mdo#S32259213	ARR-B (419669)	2.0E-15	NGS	5.33



trans-Zeatin reversibel O-glucosyliert und damit inaktiviert. Verringerte Phytohormonkonzentrationen liegen dagegen für polare Gibberelline durch die deutlich schwächere Expression von GA2ox1 vor.

In der Biosynthese von Auxin involvierte Gene sind nicht signifikant reguliert. Allerdings führt die Regulation der Mitglieder der GH3-Familie zu geringeren Mengen Aminosäure-gekoppelter IAA (IAA-AS-Konjugat, inaktiv) im SAM von P28. Auch der Transport freier IAA (nicht gekoppelt, aktiv), die dadurch in höheren Konzentrationen im Gegensatz zur normalen Wuchsform vorliegt, ist durch die erhöhte Transkription der Gene für Ein- und Ausfluss-Transporter (PIN1 und AUX1) gefördert. Über den Hormonhaushalt von Abscisinsäure (ABA), Ethylen und Strigolactonen kann anhand der Transkriptomdaten keine Aussage getroffen werden.

3.5.4 LOKALISATION DIFFERENTIELL REGULIERTER GENE

Aus den bis hierher gezeigten Ergebnissen kann keine Verknüpfung zu einem potentiellen Kandidatengen für *Co* hergestellt werden, dessen Genprodukt die Auswirkungen auf eine solch große Anzahl von Genen in der beobachteten Art und Weise erklären würde. Daher wurden in einer abschließenden Analyse alle Gene nochmals in Betracht gezogen, die sich mit mindestens zwei Methoden als signifikant differentiell reguliert nachweisen ließen (Tabelle 3.10). Dabei wurde zu den drei bislang vorgestellten Methoden DESeq, Microarray und qRT-PCR zusätzlich eine sehr stringente Kartierung aller Illumina-Daten im Genom von *Malus x domestica* [9] mit Bowtie [154] vorgenommen und dies als vierte Methode zur Studie der Genexpression deklariert (Tabelle 3.10 ‚Bowtie‘). Insgesamt ergeben sich daraus 24 Gene, von denen 9 in P28 herunter und 15 hoch reguliert sind. Fünf davon sind bereits aus der Gruppe der Phytohormon-assoziierten Gene (Tabelle 3.9) bekannt (Nr. 1, 9, 11, 15, 24), drei lassen sich mit biotischem Stress (2, 7, 8), zwei weitere mit Transport (12, 13) assoziieren – funktionelle Kategorien, die schon durch die MapMan Analyse (Abbildung 3.13) als überrepräsentiert eingestuft wurden. Betrachtet man nun die chromosomale Lage dieser 24 Gene, so fällt auf, dass sieben von ihnen auf Chromosom 10 lokalisiert sind, auf welchem sich ebenfalls der *Co*-Locus befindet. Alle weiteren Gene scheinen bezogen auf ihre chromosomale Lage auf den weiteren 16 Chromosomen normalverteilt zu sein.

Zum Überprüfen ob die Anreicherung von *Co*-Gen abhängig regulierten Genen auf Chromosom 10 signifikant ist, wurde ein χ^2 - Test für diskrete Verteilungen durchgeführt. Unter der Annahme, dass alle Gene des Apfels auf den 17 Chromosomen in etwa gleich verteilt vorliegen, lässt sich so die Nullhypothese (dass die 7 von 24 auf Chromosom 10 gefundenen Gene dennoch einer Gleichverteilung unterliegen) mit dem Signifikanzniveau $\alpha = 0,025$ ablehnen. Demnach liegt mit einer hohen Wahrscheinlichkeit eine Anhäufung von Genen mit signifikant differentieller Regulation auf Chromosom 10 vor. Dies sind die bereits kennen gelernten Gene XTH, CYP375A1, GA2ox1, SQLE und darüber hinaus Gene für Peroxidase 42 (PER42), eine Polyphenoloxidase (PPO) und eine 70kDa Peptidylprolyl Isomerase (ROF1). Die genaue Lokalisation dieser Gene auf Chromosom 10 ist in Abbildung 3.15 dargestellt.

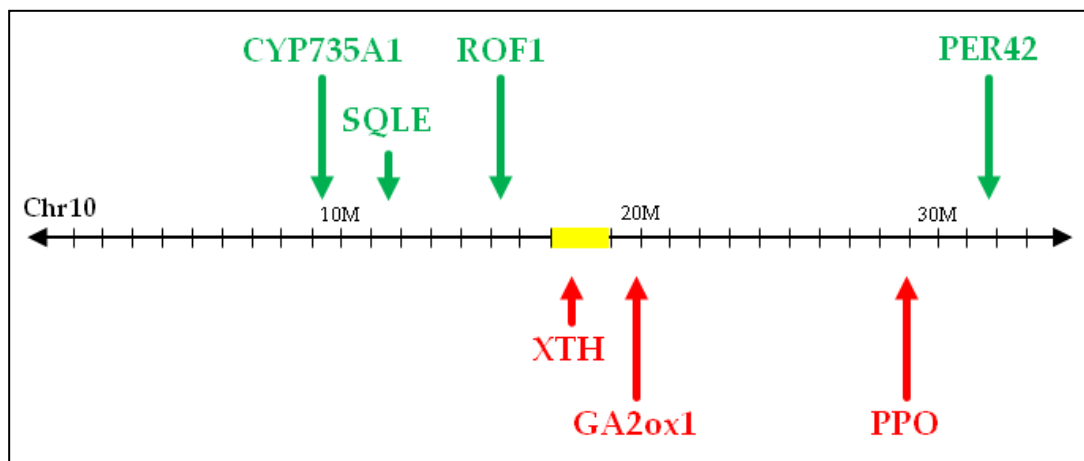



Abbildung 3.15: Anreicherung differentiell regulierter Gene auf Chromosom 10. Die Grafik zeigt die genaue Lokalisation der vier in P28 hoch (grün) und drei herunter (rot) regulierten Gene. Der potentielle Locus des *Co*-Gens ist gelb markiert. M – Megabasen.

Tabelle 3.10: Lokalisation von signifikant differentiell regulierten Genen im Genom von *Malus x domestica*. Die abschließende Tabelle beinhaltet 24 Gene, die sich mit mindestens zwei der vier Methoden zur Analyse der Genexpression als signifikant differentiell exprimiert herausgestellt haben. Angegeben ist die chromosomale Lokalisation im publizierten Genom von *Malus x domestica* [9] und die differentielle Regulation als \log_2 „change fold“ mit farblicher Unterlegung nach der unten angegebenen Farbskala. Negative Werte und rote Unterlegung stehen für herunter regulierte, positive Werte und grüne Unterlegung für in P28 hoch regulierte Gene. Horizontale Striche markieren, wo Gene nicht

vorhanden (DESeq) oder nicht differentiell reguliert (MA) waren. Chr. – Chromosom; MA – Microarray; n.b. – nicht bestimmt.

Funktion	Chr.	Bowtie	DESeq	MA	qPCR
1 Xyloglucan endotransglycosylase	10	-3.12	-2.20	-0.73	-1.20
2 Thaumatin-like protein 1b	?	-2.79	-2.59	-0.99	-1.33
3 Epidermis-specific secreted glycoprotein EP1	3	-5.16	-2.32	0.64	-2.06
4 Sieve element occlusion b	13	-2.23	-3.19	-1.53	n.b.
5 Polyphenol oxidase, chloroplastic	10	-1.14	-4.57	-1.03	n.b.
6 Splicing factor U2af large subunit A	5	-6.01	-	-	-2.17
7 Bax inhibitor 1	6	-3.12	-	n.b.	-1.14
8 Putative DnaJ protein	17	-6.17	-2.59	n.b.	n.b.
9 Gibberellin 2-beta-dioxygenase 1	10	-4.77	-2.50	n.b.	n.b.
10 Thioredoxin O1, mitochondrial	8	4.44	3.36	1.53	4.72
11 Cytokinin hydroxylase CYP735A1	10	4.97	4.10	2.50	2.78
12 Probable peptide/nitrate transporter At5g13400	13	1.55	5.30	1.07	1.20
13 Vacuolar cation/proton exchanger 3	12	2.07	5.63	1.19	0.92
14 Peroxidase 42	10	1.86	-	0.91	1.51
15 UDP-glycosyltransferase 87A1	11	3.22	-	2.40	0.77
16 Interferon-induced GTP-binding protein mx, putative	2	1.76	-	1.25	1.20
17 Unknown	1	8.10	6.89	1.85	6.00
18 Predicted protein	17	5.24	-	0.80	n.b.
19 GDSL esterase/lipase	9	2.43	-	0.83	n.b.
20 Glycerol-3-phosphate acyltransferase	8	0.85	3.86	0.77	n.b.
21 Probable inactive purple acid phosphatase 27	14	4.44	4.17	0.84	n.b.
22 High mobility group B protein	14	-0,11	-	0.82	1.31
23 70 kDa peptidyl-prolyl isomerase	10	-0.34	4.29	0.73	n.b.
24 Squalene monooxygenase	10	0,03	3.39	n.b.	0.58



Darüber hinaus scheint jedoch zwischen diesen Genen kein offensichtlicher Zusammenhang zu bestehen. So lassen sich weder die Präferenz bestimmter Loci (die Gene sind von 9 bis 32 Mb verteilt), noch bekannte Interaktionen oder gemeinsame Promotorelemente nachweisen. Interessant ist aber in jedem Fall, dass zumindest drei in die Biosynthese von Hormonen involvierte Schlüsselenzyme (CYP735A1, GA2ox1 und SQLE) in nächster Nähe zum im Fokus stehenden Locus zu finden sind und eine signifikante, differentielle Regulation aufweisen, und dies unabhängig von Methode, Entnahmezeitpunkt und Material (*in vitro* und *in vivo*).

3.6 CHROMOSOMENPRÄPARATE UND FISH

Da zu Beginn der Arbeit noch keine geeignete Referenz in Form eines sequenzierten Genoms zur Verfügung stand, sollte ein FISH-System etabliert werden, um BAC-Klone (Doktorarbeit Dominik Otto [103]) und interessante Kandidatengene auf chromosomaler Ebene lokalisieren zu können.

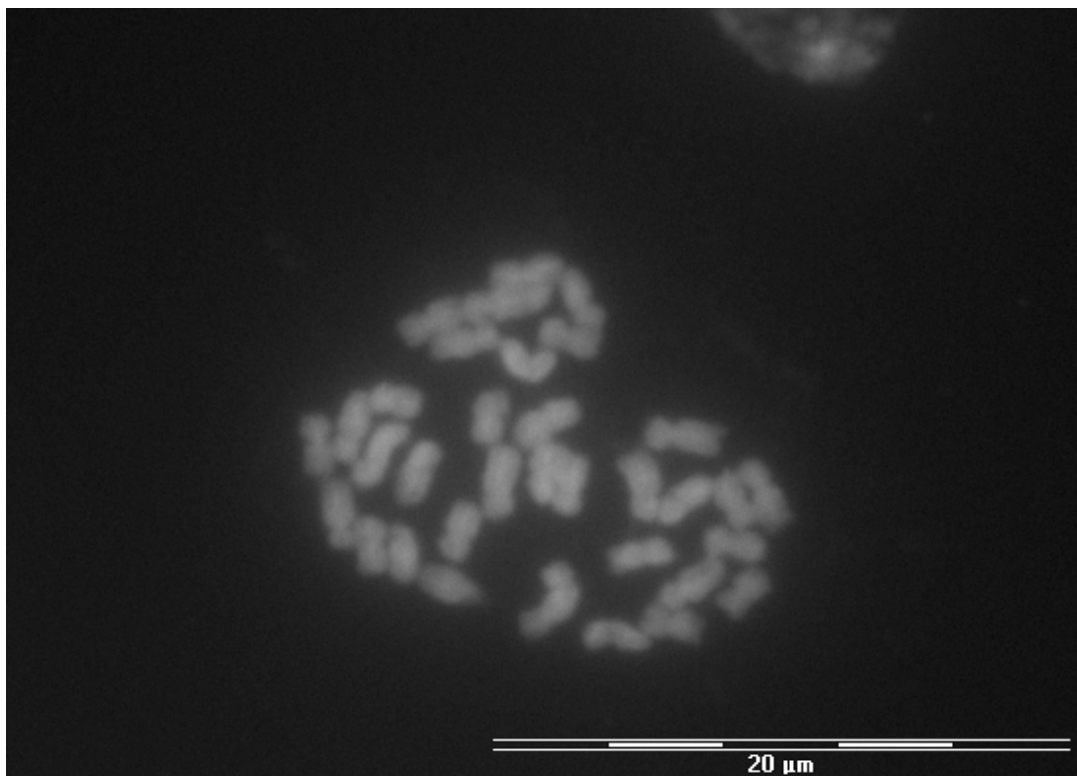


Abbildung 3.16: Metaphasechromosomen. Die Aufnahme zeigt einen vollständigen, diploiden Chromosomensatz ($2n = 34$) von *Malus x domestica* zum Zeitpunkt der Metaphase (1000-fache Vergrößerung). Am oberen Bildrand ist ein Teil des Zellkerns einer in der Interphase befindlichen Zelle zu sehen. Die Färbung der DNA erfolgte mit DAPI.

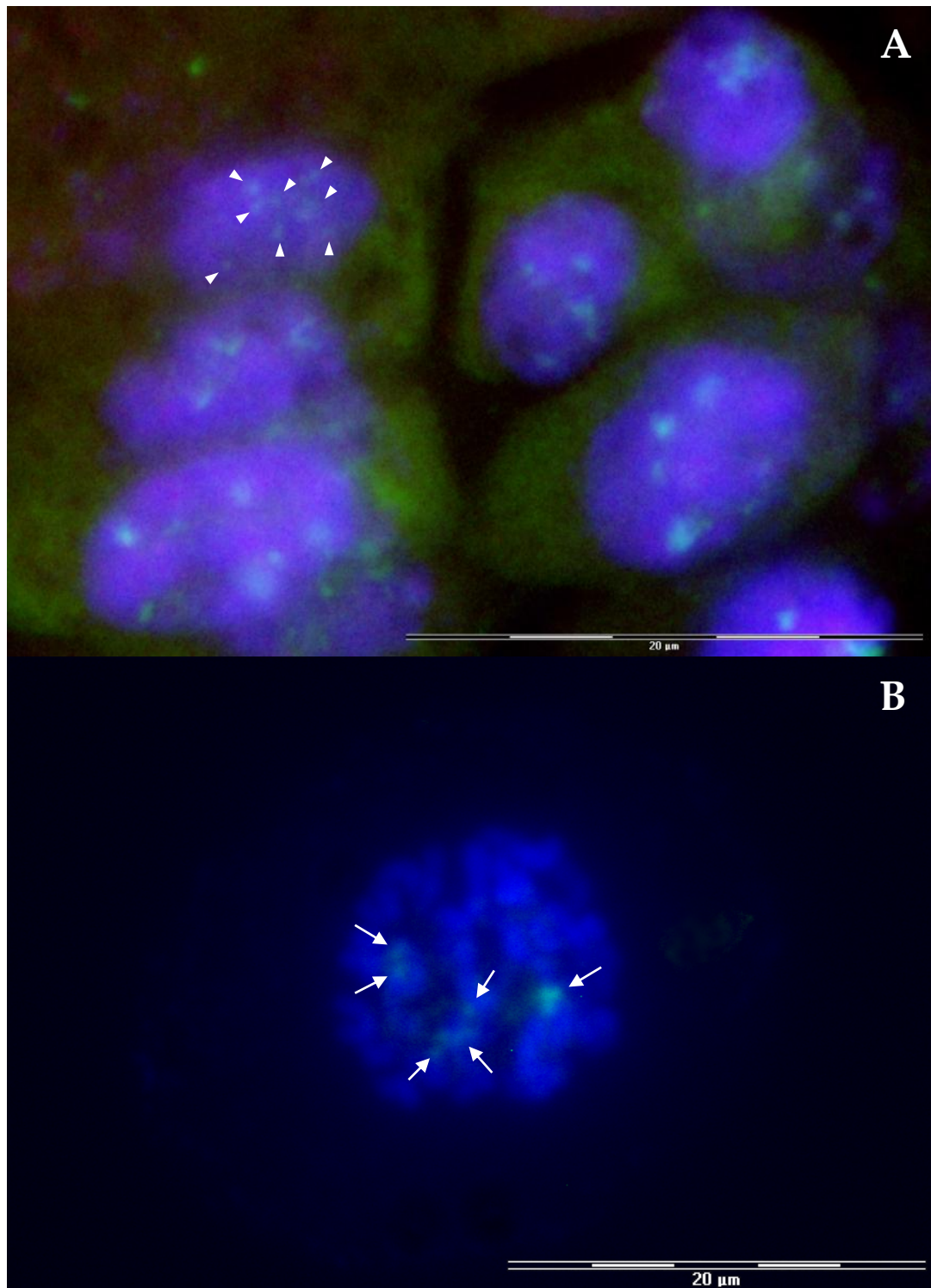


Abbildung 3.17: Nachweis von rDNA. Aufnahmen von Interphasezellen (A) und einem Prometaphasezellkern (B) in 1000facher Vergrößerung nach FISH mit einer 18S/28S rDNA-spezifischen Sonde. Die grüne Färbung zeigt die Loci der Gene für die rRNA. Dies sind bis zu acht (weiße Pfeilspitzen) in den Interphasezellen (A), während im Prometaphasezellkern (B) sechs Loci eindeutig identifiziert werden können (weiße Pfeile). Gegengefärbt wurden die Präparate mit dem DNA-spezifischen Fluoreszenzfarbstoff DAPI (blau).

Der erste Schritt dieses Vorhabens war daher die Generierung von Chromosomenpräparaten, an denen sich eine FISH aussichtsreich durchführen lassen würde. Aufgrund der hohen Zellteilungsrate wurde meristematisches Gewebe der Wurzel junger Keimlinge benutzt. Die Grundlage des finalen Protokolls wurde aus mehreren Referenzen zusammengestellt, die bereits erfolgreich FISH an pflanzlichen Chromosomen etabliert hatten [140, 176-178]. Das Ergebnis einer erfolgreichen Präparation (Arretierung der Zellteilung in der Metaphase, Protoplastenisolation und Spreiten der Chromosomen) zeigt Abbildung 3.16. Darin kann man erkennen, dass die im voll kondensierten Zustand etwa 1-3 μm langen Chromosomen überwiegend meta- bzw. submetazentrisch sind. Der hier gezeigte, diploide Satz hat $2n = 34$ Chromosomen.

Die eigentliche FISH wurde zunächst mit Sonden für Teile der 18S/28S rDNA etabliert, da die Gene für diese rRNAs hochgradig repetitiv sind und somit leichter nachzuweisen sind als eine single copy Sonde, wie sie beispielsweise ein BAC-Klon darstellt. Des Weiteren wurde in einem vergleichbaren Ansatz [179] bereits nachgewiesen, dass eine ähnliche Sonde an acht Loci auf verschiedenen Chromosomen von *Malus x domestica* hybridisiert. Mit der in dieser Arbeit generierten Sonde ließen sich diese Erkenntnisse sehr gut reproduzieren (Abbildung 3.17), auch wenn sich die Zahl von acht Loci nur in Interphasezellkernen bestätigte (Abbildung 3.17 A). Da gut geeignete Metaphasechromosomen oft nur schwache oder gar keine Fluoreszenz zeigten, konnten hier nur maximal sechs eindeutig voneinander differenzierbare Loci detektiert werden (Abbildung 3.17 B).

Im Folgenden wurde dieses System auf BAC-Klone mit 40 bis 80 kbp Integrat angewendet [103], jedoch konnte nur für den größten verfügbaren Klon (K63, 80 kbp) eine Hybridisierung nachgewiesen werden. Da die Sonde sich nur in kondensierten Chromosomen und nicht in Interphasezellkernen detektieren ließ, wurde versucht, das Signal mit geeigneten Methoden zu verstärken. Dazu gehörte der Einsatz sekundärer, ebenfalls Fluoreszenz-gekoppelter Antikörper sowie der Gebrauch von Tyramid Signalamplifikationssystemen (TSA). Leider führte dies nicht zum gewünschten Erfolg und die Reproduktion dieser einmaligen Hybridisierung ließ sich nicht bewerkstelligen.

Mit Veröffentlichung des sequenzierten Genoms von *Malus x domestica* cv. 'Golden Delicious' [9] rückte die FISH zur Lokation von BAC-Klonen und Kandidatengenen in den Hintergrund, da dies nun auf Sequenzebene mit weitaus geringerem Aufwand ermöglicht wurde.

4 DISKUSSION

4.1 EINFLUSS VON MATERIALENTNAHME UND -VERARBEITUNG

Sequenzanalysen sind gerade bei RNA in besonderer Weise vom Ausgangsmaterial abhängig. Bei der Entnahme und Weiterverarbeitung des Freilandmaterials wurde deshalb größte Sorgfalt darauf gelegt, etwaige Veränderungen in der Genexpression durch das unvermeidbare Hinzufügen eines abiotischen Stimulus [163, 164] zu verhindern. Dies wurde durch unmittelbare Kühlung der Probe und rasches Einfrieren bei -80°C realisiert. Des Weiteren wurden die Sprossspitzen weit unterhalb des SAM abgetrennt, um eine Beeinflussung dieses Gewebes weitgehend auszuschließen und die aktuelle Genexpression zum jeweiligen Entnahmezeitpunkt bestmöglich zu konservieren. Bei der Extraktion der RNA aus dem tiefgefrorenen Gewebe wurde ebenfalls zügig gearbeitet, um eine mögliche Degradation durch das Auftauen des Gewebes [180] zu vermeiden, was unweigerlich zu Beeinflussungen in der Messung der Genexpression führt [181]. Der Fokus lag hierbei klar auf einem vertretbaren Kompromiss zwischen der angesprochenen Geschwindigkeit und dennoch sorgfältiger Präparation, um das umliegende Gewebe weitgehend zu entfernen. Für die Gewinnung der RNA eines Entnahmezeitpunktes wurden stets mehrere (i.d.R. zwischen 10 und 15) freipräparierte SAM von unterschiedlichen Bäumen (gleiche Kreuzung, gleicher Standort) benutzt, um einerseits eine gewisse Menge an Ausgangs-RNA für die Folgeanwendungen zu garantieren und andererseits individuelle Unterschiede zwischen einzelnen Bäumen weitgehend zu kompensieren. Die gefundenen Unterschiede sollten daher nur auf Abweichungen im Vergleich kolumnarer und nicht-kolumnarer SAM basieren. Auch für die qRT-PCR wurden mehrere *in vitro*-Kulturen einer Charge benutzt. Aufgrund der geringen Größe (ca. 2-3 cm) wurde hier auf eine weitere Präparation – abgesehen von der Entfernung des Kallus – verzichtet. Der Probenpool sollte weiterhin dazu dienen, Abweichungen einzelner Proben zu puffern, deren Genexpression

beispielsweise durch nicht entfernte seneszente Bereiche (*ivK*) oder nicht offensichtliche Schäden durch Pathogenbefall (Freilandmaterial) beeinträchtigt ist. Für alle Folgeanwendungen mit der so gewonnenen Total-RNA als Ausgangsmaterial (qRT-PCR und Microarrays) wurden für die Messung in A14 und P28 vergleichbare RIN-Werte [136] größer 8,0 vorausgesetzt, da degradierte RNA (kleine RIN-Werte) und große Abweichungen zwischen den RIN-Werten verschiedener RNAs die Ergebnisse maßgeblich beeinflussen [135, 182]. Bei den RNAs für die Transkriptomsequenzierungen wurden dagegen weniger strenge Qualitätskriterien angelegt, da hier eine große Menge Total-RNA (ca. 100 µg) für die Poly-A⁺-Aufreinigung benötigt wurde. Alle weiteren Proben wurden demnach zwecks besserer Vergleichbarkeit ebenfalls linear amplifiziert. Die Verwendung dieser Methode legt den Schluss nahe, dass gerade Studien zur Genexpression beispielsweise durch Bevorzugung bestimmter Transkripte oder unvollständige Amplifikation beeinflusst werden. Dies konnte jedoch in diversen vergleichenden Genexpressionsstudien mit Hilfe von qRT-PCR und Microarrays anhand von linear amplifizierter und nicht amplifizierter RNA widerlegt werden [183-186]. Auch in dieser Arbeit konnte gezeigt werden, dass die auf nicht amplifizierter RNA basierenden qRT-PCR- und Microarray-Ergebnisse gut mit den NGS Daten (auf der Basis von amplifizierter aRNA) übereinstimmen. Einen Vorteil der aRNA gegenüber der Verwendung von mRNA konnte durch den Vergleich der RNAs in Abbildung 3.1 deutlich gemacht werden: im Vergleich zu 9,7 bzw. 6,5 % rRNA-Kontamination in der mRNA ist durch die zweite Aufreinigungsrunde mit oligo(dT) Primern in der aRNA keine Kontamination mehr nachzuweisen, was durch die Sequenzierungen bestätigt wird. Dies führt aber ebenfalls zu der starken 3'-Verschiebung der Sequenzen (Abbildung 3.3), die zwar keinen nennenswerten Einfluss auf die hier durchgeführten Genexpressionsstudien haben sollte, die vollständige Analyse der Gene auf Nukleotidebene aber vor allem im 5'-Bereich erschwert.

4.2 NGS UND DATENVERARBEITUNG

Für den vollständigen Vergleich der Transkriptome von A14 und P28 wurden in dieser Arbeit sowohl Roche 454- als auch Illumina-Technologie eingesetzt. Die generellen Vor- und Nachteile dieser Sequenzierstrategien für verschiedene Anwendungen wurden bereits eingehend beleuchtet [113, 187, 188], daher soll an

dieser Stelle nicht detailliert darauf eingegangen werden. Alleine die quantitative Ausbeute an Sequenzinformation ist bei in etwa vergleichbaren Kosten mit Illumina 80-fach höher, allerdings bei doppelter Fehlerrate (Tabelle 3.2 und Tabelle 3.3). Dies und die höhere Abdeckung bekannter Transkripte (Abbildung 3.3) favorisieren klar die Illumina-Technologie für RNA-Seq Experimente. Die kurzen Leseweiten erschweren jedoch Kartierungen oder BLAST-Suchen, insbesondere wenn keine geeignete Referenz zur Verfügung steht.

Beim Vergleich der Assemblierungen zeigt sich dagegen der Vorteil längerer Leseweiten, da die beiden Roche 454-Datensätze erwartungsgemäß in annähernd gleiche Zahlen von Contigs assembliert werden. Diese stellen mit ca. 24.000 (Tabelle 3.5) eine durchaus realistische Anzahl bezogen auf die zu diesem Zeitpunkt exprimierten Gene dar. Ein Vergleichswert existiert leider nicht, da in diesem Umfang bislang keine Expressionsstudien beim Apfel durchgeführt wurden. Dagegen gestaltet sich die *de novo*-Assemblierung der Illumina-Daten schwierig, wenn die Zahl der zum jeweiligen Zeitpunkt exprimierten Gene gänzlich unbekannt ist. Hier liefert der ABySS-Algorithmus mit nur leicht abweichenden Parametern extrem unterschiedliche Ergebnisse. Eine Anpassung der Parameter ist bei *de novo*-Assemblierungen jedoch obligatorisch [165], daher empfiehlt sich bei solchen Vorhaben die Kombination beider Methoden, da man die Parameter soweit anpassen kann, bis näherungsweise die verlässlicheren Ergebnisse reproduziert werden. Eine weitere Möglichkeit besteht darin, die Assemblierung der kurzen „reads“, sofern vorhanden, an einer geeigneten Referenz durchzuführen. *De novo* konnten die Illumina-Sequenzdaten im bestmöglichen Fall zu 30.643 (A14) bzw. 39.194 (P28) assembliert werden, wobei der Unterschied von 8.500 Contigs auf den Überlappungsgrad der „paired-end reads“ zurückzuführen ist, der in P28 höher ist (Tabelle 3.4) und damit zu kürzeren Ausgangssequenzen führt. Die Zahl der exprimierten Gene zu diesem Zeitpunkt kann zwar nicht exakt beziffert werden, da abhängig von den jeweilig gewählten Parametern beispielsweise auch Isoformen oder allelische Varianten zu eigenen Contigs assembliert werden und somit die Zahl variiert. Der Wahrheitsgehalt beider Assemblierungen ist dennoch als hoch einzuschätzen, da unabhängig von der Abdeckung des gegenwärtigen Transkriptoms die gleiche Zahl an nichtredundanten Basen zu Contigs assembliert werden. Dies sind für alle vier Datensätze ca. 18 Mb, wodurch von einer kompletten Transkriptomabdeckung sowohl der Roche 454- (ca. 5-fach) als auch der Illumina-Daten (ca. 80-fach) ausgegangen werden kann. Die geringere Abdeckung durch die Roche 454-Daten hat ein größeres Fehlerpotential zur Folge, da teilweise Sequenzen

nur durch wenige Sequenzen abgedeckt sind. Auf der anderen Seite erhält man insgesamt mehr längere Contigs, die einem höheren Informationsgehalt (gemessen an der Transkriptlänge) entsprechen. Auch hier bietet sich die Kombination beider Technologien an, um zunächst weniger, dafür aber längere Contigs zu generieren, um deren Sequenz anschließend mit den Illumina-Daten zu bestätigen, SNPs und damit Allele oder verschiedene Isoformen zu identifizieren. Dies wurde in der vorliegenden Arbeit nur für einige interessante Gene durchgeführt, da das Hauptaugenmerk auf Unterschieden in der Expression, nicht im Auffinden von Genvarianten lag.

Darüber hinaus liefert die Auswertung mit der in dieser Arbeit entwickelten und beschriebenen Pipeline (Abbildung 3.6) genauere und vergleichbarere Ergebnisse, wenn nicht assemblierte, sondern Rohdaten als Eingabe verwendet werden. Dies wird zum einen an der Verteilung der Spezies auf der Basis von 500.000 BLAST-Treffern (Abbildung 3.7) deutlich. Diese liefert für zwei verschiedene Datensätze (unterschiedliche Entnahmezeitpunkte und Sorten) annähernd identische Ergebnisse, während Speziesverteilungen basierend auf Contigs der gleichen Datensätze (nach Assemblierung) wesentlich stärker divergieren (nicht gezeigt). Um die gezeigte Verteilung und das darauf basierende Kladogramm (Abbildung 3.8) nun aber sinnvoll bewerten zu können, muss die Anzahl der in der Datenbank (nt) enthaltenen Sequenzen und deren relativer Anteil mit berücksichtigt werden [SI Tabelle_2]. So wird an der Zahl der Treffer der hohe Verwandtschaftsgrad zwischen *Malus* und *Pyrus* deutlich, die beide dem Tribus Maleae zugeordnet werden können. Noch mehr Gene als mit dem nächsten hier vertretenen Verwandten, *Prunus* (alle Subfamilie Maloideae) teilt sich der Apfel mit *Ricinus*, gefolgt von *Prunus*, *Vitis* und *Populus*. *Lotus*, *Glycine*, *Solanum* und *Arabidopsis* zeigen gemessen an den in der Datenbank abgelegten ESTs (75 % der zehn hier aufgeführten Gattungen) nur geringe Übereinstimmungen. Dies verdeutlicht, dass die meisten der im Sprossapikalmeristemgewebe von *Malus x domestica* exprimierten Gene auch in anderen stark verholzten Pflanzen der Subklasse Rosidae (und weiteren, zusammengefasst in ‚Andere‘, Abbildung 3.7) vorkommen. Sträucher und krautige Pflanzen wie *Arabidopsis* und *Glycine*, die aufgrund ihres Generationszyklus bessere Modellorganismen darstellen und somit weitaus mehr Datenbankeinträge haben, sind gemessen an der Zahl letzterer eindeutig unterrepräsentiert. Selbst die Weinrebe (*Vitis vinifera*) stellt trotz eines schwächeren Verwandtschaftsverhältnisses einen weitaus höheren relativen Anteil dar als die beiden letztgenannten Gattungen. An dieser Stelle lassen sich aber noch keine Aussagen über etwaige Unterschiede

der beiden Sorten A14 und P28 ausmachen, da die minimale Abweichung der Stichprobe statistisch nicht signifikant ist.

Den Vorteil der Verwendung von Rohdaten im Hinblick auf differentielle Regulation zeigt Abbildung 3.9, in der dieser Vergleich anhand unterschiedlich weiterbearbeiteter Datensätze, die auf exakt demselben Ausgangsmaterial (aRNA) beruhen, visualisiert ist. Alle Fehlerquellen können auch in dieser endgültig verwendeten Variante nicht ausgeschlossen werden. Vor allem durch die Verwendung des BLAST-Algorithmus in Kombination mit den kurzen Suchsequenzen (100 Basen) ist nicht auszuschließen, dass zwei „reads“ desselben Transkriptes aufgrund geringfügiger Abweichungen in der Sequenz (z.B. bedingt durch falsches Basecalling) zwei unterschiedliche Treffer generieren. Dieser Gefahr kann lediglich durch die Wahl einer geeigneten Datenbank entgegengewirkt werden. Zwar ergeben sich bei größeren Datenbanken (z.B. RefSeq) mehr Einträge als bei kleinen, gut annotierten (z.B. UniProt), die Qualität der Treffer ist bei abnehmendem Informationsgehalt jedoch in der Regel höher. Erwartungsgemäß liefert die Verwendung speziesspezifischer Datenbanken (in diesem Falle MDPs oder *Malus* Unigene) die verlässlichsten Ergebnisse, ist jedoch auf Spezies limitiert, für die eine solche existiert. Benötigt man hingegen möglichst genaue Funktionsinformationen (z.B. für Gene Ontology Analysen), bietet sich eine zweite, hochstringente BLASTx- oder BLASTp-Suche der in der ersten BLAST-Suche erhaltenen Treffer gegen eine entsprechende Datenbank an. Für diesen Fall wurden die Skripten der Pipeline so programmiert, dass sehr ähnliche Treffer (dasselbe Gen in nahe verwandten Spezies) auf einen Eintrag (entsprechend mit zwei Treffern) kollabiert werden.

Mit der hier etablierten Pipeline können also nahezu beliebig große Datenmengen analysiert werden, da sowohl BLAST-Suchen wie auch Kartierungen hochparallel und dadurch sehr schnell auf HPC-Clustern durchgeführt werden können. Eine aufwendige, fehleranfällige und zeitintensive Assemblierung kann somit übergangen werden. Die Pipeline erlaubt den Einsatz jedweder Datenbank und ist damit nicht in der Wahl des Organismus oder der Anwendung limitiert. Darüber hinaus kann an dieser Stelle auch jeder beliebige Filter angewendet werden. Abhängig von der jeweiligen Referenz kann man sich zwischen BLAST-Suchen und hochstringenten Kartierungen entscheiden und dabei alle verfügbaren Parameter im Vorfeld der Suche anpassen. Die Resultate werden anschließend bei Genexpressionsstudien mit Hilfe von BORT automatisch quantifiziert und

anwenderbestimmt ausgegeben und können so weiterreichenden Anwendungen beziehungsweise Folgeanalysen zugeführt werden.

4.3 HOHE ANZAHL DIFFERENTIELL REGULIERTER GENE IM SAM KOLUMNARER BÄUME

Wie eingangs bereits erwähnt, sollten Gene identifiziert werden, die unabhängig von Entnahmezeitpunkt und Methode im SAM kolumnarer Bäume eine differentielle Regulation zeigen, worunter auch solche fallen, die aufgrund der Verwendung unterschiedlicher Sorten abweichend exprimiert werden. Diese Gene können von denen im Interesse stehenden durch die hier verwendeten Methoden nicht eindeutig separiert werden. Allerdings liegt die Vermutung nahe, dass diese Zahl eine untergeordnete Rolle spielt, da sich die Phänotypen von A14 und nicht-kolumnarer P28 in hohem Maße ähnlich sind. Der Großteil an differentiiell regulierten Genen sollte demnach tatsächlich auf die Anwesenheit des *Co*-Allels zurückzuführen sein.

Der eingehende Vergleich und die Analyse differentiiell regulierter Gene zweier Zeitpunkte ist bereits Gegenstand einer im Zuge dieser Arbeit entstandenen Publikation [189], daher wird hier ein darauf aufbauender Vergleich unter Einbeziehung sämtlicher Illumina-Datensätze gewagt. Die Gegenüberstellung der Pearson Koeffizienten (Abbildung 3.10) zeigt durch Werte von 0,69 bis 0,86 generell starke Homologien sowohl zwischen A14 und P28 wie auch zwischen den verschiedenen Zeitpunkten auf. Dies ist für die geplanten Analysen äußerst günstig, da dies auf nur wenige, sortenspezifisch exprimierte Gene hinweist. Da selbst bei technischen Replikaten keine Werte von 1,0 erreicht werden [190], geht die Abweichung zu absoluten Homologien zwischen A14 und P28 auf differentiiell regulierte Gene zurück, während die Abweichungen der unterschiedlichen Entnahmezeitpunkte innerhalb einer Sorte auf andere Expressionsniveaus der entsprechenden Gene und/oder die Expression verschiedener Gene hindeuten. Dies lässt sich durch die Streudiagramme (Abbildung 3.11) und die Betrachtung der am höchsten exprimierten Gene verschiedener Zeitpunkte (Tabelle 3.7) bestätigen.

Die einzige Abweichung stellt der Datensatz A14 07-2010 dar, welcher große Differenzen zu allen Datensätzen bis auf P28 07-2010 zeigt. Ein technischer Fehler bei der Erstellung der Bibliothek oder der Sequenzierung kann hier nicht gänzlich ausgeschlossen werden, da die verwendete aRNA Anzeichen einsetzender

Degradation aufwies. Die Analyse des Datensatzes im Vergleich zu anderen zeigt jedoch keine offensichtlichen Unterschiede. Vielmehr handelt es sich hierbei um ein bereits in kolumnaren Apfelsorten beobachtetes Phänomen: während die vegetative Wachstumsperiode normalwüchsige Bäume von in etwa Mai bis Juli dauert, wachsen kolumnare Bäume bis in den Oktober hinein [105, 191]. Da das Material am 20. Juli entnommen wurde, könnte es also durchaus sein, dass A14 das Wachstum bereits eingestellt hatte, was mit der beobachteten, massiven Veränderung der Genexpression in A14, nicht aber in P28, einhergehen würde. Eine weitere Erklärung wäre, dass A14 und P28 unterschiedlich auf supraoptimale Bedingungen reagieren, da zum Zeitpunkt der Entnahme Temperaturen um 34°C vorherrschten. Diese Annahmen sollten jedoch durch weitere Sequenzierungen des entsprechenden Zeitpunktes eingehender geprüft werden.

Ungeachtet dessen zeigt der Vergleich der NGS Daten von A14 und P28 zu allen Zeitpunkten eine sehr hohe Anzahl von differentiell exprimierten Genen, sowohl ausgehend von Rohdaten als auch von normalisierten „read counts“ (RPKM). Sucht man jedoch gezielt nach Genen, die zu mehreren Zeitpunkten in gleichem Maße differentiell reguliert sind (Abbildung 3.12 und Abbildung 3.13), verringert sich diese Zahl auf einige hundert. Neben den schon thematisierten, Entnahmezeitpunkt abhängig differentiell regulierten Genen (Kat. II und IV, Abbildung 3.12) finden sich mehr Gene in P28 hoch als herunter reguliert, was durch die Ausweitung auf alle sechs Datensätze (DESeq) bestätigt wird. Anhand dieser Erkenntnisse lässt sich daher eher auf eine regulatorisch aktivierende als inhibierende Wirkung des Co-Genproduktes schließen.

4.4 DIFFERENTIELL REGULIERTE GENE UND KOLUMNARER PHÄNOTYP

Die funktionelle Analyse der unabhängig vom Entnahmezeitpunkt differentiell regulierten Gene wurde auf Grundlage der DESeq-Ergebnisse durchgeführt, da diese durch die zu Grunde liegende Datenmenge als statistisch abgesichert zu bewerten sind. Gene mit $\text{baseMean} \leq 150$ wurden von der Analyse ausgeschlossen, da geringere Werte eher basaler Transkription ($\text{RPKM} \leq 10$) entsprechen und deren Wirkung daher als gering einzustufen ist. Die verbleibenden 619 Gene mit $p < 0,005$ und „fold change“ $> 3,0$ zeigen bei der Analyse mit MapMan neun überrepräsentierte, funktionale Kategorien (BINs), in die sich 147 der 465

gruppierten Gene einordnen lassen (31,6 %). Diese Ergebnisse bestätigen zumindest sechs BINs, die bereits vorher als überrepräsentiert identifiziert wurden und deren Gene sich aufgrund ihrer Regulation mit dem kolumnaren Phänotyp assoziieren lassen [189].

Die zahlreichen Gene in den BINs ‚Transport‘ und ‚Zellwanddegradation‘ lassen auf veränderte Membran- und Zellwandzusammensetzungen schließen. Von vielen dieser Gene ist bekannt, dass eine starke Hoch- oder Herunterregulation zu gravierenden phänotypischen Abweichungen führt [192, 193], was teilweise in Einklang mit dem vorliegenden Habitus gebracht werden kann, diesem teilweise aber auch gänzlich widerspricht. Alleine anhand der Expressionsstudien kann ohnehin keine Aussage getroffen werden, ob die geänderte Expression dieser Gene den Phänotyp bedingen können. Es ist jedoch wahrscheinlich, dass die kürzeren Internodien und der größere Stammdurchmesser auf Modifikationen dieser funktionellen Gruppen zurückzuführen sind.

Ebenfalls interessant sind die fast ausschließlich in P28 hoch regulierten Gene, die sich den BINs ‚Proteinmodifikation und Degradation‘ zuordnen lassen. In Eukaryoten ist der Abbau spezifischer Proteine für Entwicklungsprozesse essentiell und daher sehr genau reguliert. Dies wird über das 26S Proteasom [194] realisiert, an dessen Integrität die hier genannten Genprodukte beteiligt sind. In Pflanzen ist dieses System eng gekoppelt mit dem Hormonhaushalt (vergleiche Abbildung 1.1 **B**, Abbildung 1.3 **B**, Abbildung 1.4 **B**, Abbildung 1.5 **B**) und Reaktionen auf biotischen Stress [195] – zwei Kategorien von Genen, die ebenfalls überrepräsentiert sind und deren Vertreter auffällige Änderungen in ihrer Expression zeigen. Wie diese Regulation im Einzelnen aussieht und was der maßgebliche Faktor ist, konnte anhand der vorliegenden Daten nicht geklärt werden, ein Zusammenhang zwischen den hier genannten BINs scheint jedoch offensichtlich. Ob die Änderungen im Hormonhaushalt (vgl. 4.5) ein Resultat dieser Regulation oder im Umkehrschluss dafür verantwortlich ist, kann ebenfalls nur spekuliert werden. Allgemein scheint eine zentrale Rolle jedoch wahrscheinlich, was die Betrachtung der signifikant differentiell regulierten Transkriptionsfaktoren (BIN ‚Regulation of Transcription‘) verdeutlicht (Abbildung 3.14). Neben den offensichtlich im Hormonhaushalt involvierten Aux/IAA [196], ARR-B und Pseudo ARR [197, 198] TFs, sind auch Vertreter der GeBP-Familie (assoziiert mit der Regulation von CK-Antworten [199]) und der AP2/EREBP-Familie (herunter reguliert durch BR und CK [200]) beteiligt. Ebenfalls interessant erscheinen die Mitglieder der YABBY-Familie, die Einfluss auf

die Entwicklung des SAMs und die Verteilung des Auxin Efflux Carriers PIN1 haben [201]. Neben diesen lassen sich weitere Gene aufgrund ihrer Funktion und Regulation mit dem Kolumnarwuchs in Verbindung bringen. Dennoch lassen sich keine Faktoren benennen, deren Regulation allein den Phänotyp erklären würde. Die funktionelle Charakterisierung liefert zwar überrepräsentierte, funktionelle Kategorien, die sich teilweise miteinander assoziieren lassen, nicht aber die Interaktionen aufklären, die der hier gefundenen Regulation zu Grunde liegen. Die funktionelle Analyse für sich genommen reicht demnach nicht aus, auf eine potentielle Funktion des Co-Genproduktes zu schließen.

4.5 PHYTOHORMONHAUSHALT DES KOLUMNAREN SAMs

Es liegt die Vermutung nahe, dass Phytohormone eine wichtige Rolle bei der Etablierung des kolumnaren Phänotyps spielen. Mehrere Studien konnten auf diesem Gebiet bereits zeigen, dass Hormone wie Auxin, CK, ABA und GA in kolumnaren Bäumen unterschiedliche Konzentrationen aufweisen (Abbildung 1.9 und Text 1.3).

In dieser Arbeit konnten mit den beschriebenen Methoden insgesamt 26 in ihrer Regulation auffällige Gene identifiziert werden, die sich direkt mit dem Phytohormonhaushalt assoziieren lassen (Tabelle 3.9). Davon sind 11 am Anabolismus beteiligt (Abbildung 4.1), wodurch Schlüsse auf die endogene Konzentration des jeweiligen Hormons möglich sind, auch wenn diese nicht direkt gemessen werden konnten.

So sind SQLE und DIM/DWF1 Schlüsselenzyme in der Biosynthese von BRs (Abbildung 4.1 A) und damit direkt an der Regulation der Zellstreckung beteiligt [202], denn Funktionsverlustmutanten zeigen gravierende Defekte in Spross- und Wurzelelongation [203]. Da beide jedoch in P28 hoch reguliert sind, liegt demnach eine höhere BR-Konzentration vor. Im Umkehrschluss würde dies zu einem stärkeren Streckungswachstum führen und damit direkt an der Ausprägung des Phänotyps beteiligt sein, wenn die vorliegenden BR-Konzentrationen nicht bereits supraoptimal sind. In *Arabidopsis* führt die Überexpression von DWF1 zu phänotypisch unauffälligen, aber toleranteren Pflanzen gegenüber biotischem Stress [204]. Dies stellt eine Verbindung und gleichzeitig eine mögliche Erklärung für die in P28 herunter regulierten Gene dar, die sich dem überrepräsentierten BIN ‚Stress biotisch‘ (Abbildung 3.13 B) zuordnen lassen.

Weitere vier in P28 hoch regulierte Gene LOX, AOS, AOC4 und JAR1 sind an der Bildung von JAs und insbesondere der Synthese des hochaktiven JA-Ile [74, 205, 206] beteiligt (Abbildung 4.1 B). Dies führt zu höheren Konzentrationen eines eher wachstumshemmenden Hormons [72, 207], was der vermeintlichen Wirkung der BRs entgegenwirken würde. Es ist jedoch auch bekannt, dass BRs die Biosynthese von JAs positiv regulieren [30]. Aufgrund der komplexen Interaktionsnetzwerke verschiedener Phytohormone lässt sich auch in diesem Fall nicht eindeutig sagen, welches das bestimmende Prinzip ist. Aufgrund der vorgestellten Zusammenhänge wäre es aber durchaus denkbar, dass die BR-Synthese direkt vom Co-Genprodukt beeinflusst wird und die höhere Konzentration von BRs zur Aktivierung der JA-Biosynthese und Drosselung der Stressantwort führt.

CYP35A1 und ZOG1 sind Schlüsselenzyme in der späten Synthese des hochaktiven CKs trans-Zeatin (tZ) und dessen reversibler Inaktivierung durch O-Glucosylierung (Abbildung 4.1 C) [46, 175]. Bei der funktionellen Betrachtung von ZOG1 muss allerdings berücksichtigt werden, dass für das hier gefundene Gen (MDP000027657) zwar starke Homologien zu ZOG1 aus *Phaseolus lunatus* gefunden wurden, in der weitreichenden Familie der UDP-Glucosyltransferasen aber andere Funktionen nicht auszuschließen sind. Auf jeden Fall führt die Hochregulierung von CYP735A1 zu erhöhten CK-Konzentrationen im SAM von kolumnaren Bäumen, was frühere Messungen an ‚McIntosh Wijcik‘ bestätigen [92]. Anhand dieser Erkenntnisse lassen sich zwei Hypothesen zur erhöhten CK-Toleranz kolumnarer Apfelbäume aufstellen. Zum einen könnte der durch Co bedingte, endogen erhöhte CK-Spiegel (CYP735A1) die Resistenz gegen exogen appliziertes CK [109, 208] erhöhen, weil die Zellen sich diesen Konzentrationen angepasst haben. Auf der anderen Seite könnte auch das hoch regulierte ZOG1 dafür sorgen, dass synthetisches und natürliches CK inaktiviert wird, dies für den hohen endogenen Spiegel sorgt und gleichzeitig die hohe Toleranz erklärt. Ein solcher Faktor zur Metabolisierung überschüssigen CKs wurde bereits prognostiziert [209]. Überexpression von ZOG1 führt in *Arabidopsis* zu verzögertem Wuchs [210], was in kolumnaren Äpfeln nicht geschieht. Wie bereits gezeigt sind hier jedoch die endogenen CK-Konzentrationen generell höher und auch die Mitwirkung anderer Hormone, die abweichende Konzentrationen zu normalwüchsigen Bäumen zeigen, dürfen nicht außer Acht gelassen werden.

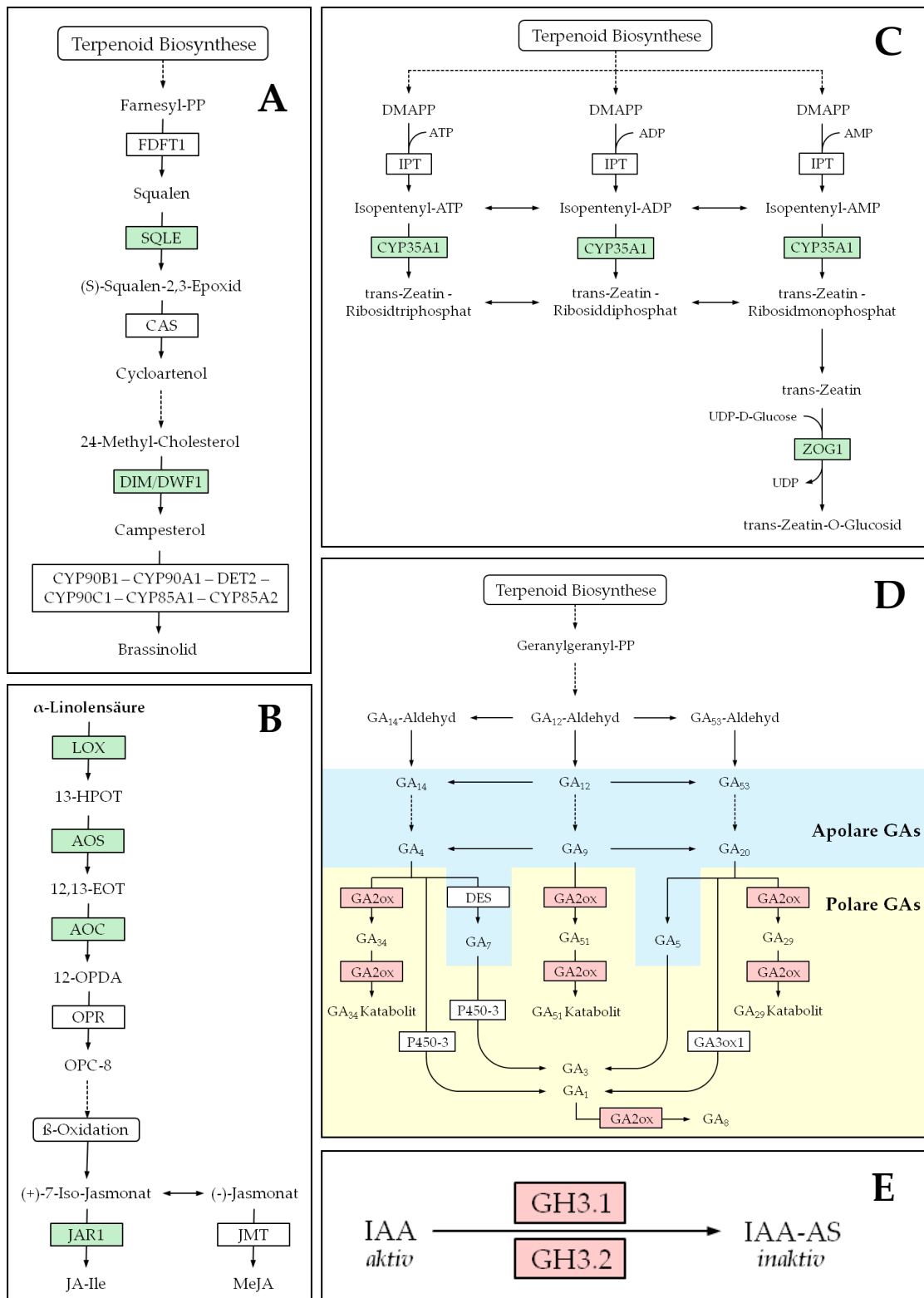


Abbildung 4.1: Hormonale Situation im SAM kolumnarer Bäume. Gezeigt sind alle signifikant regulierten, mit der Biosynthese von Phytohormonen assoziierten Gene und ihr Einfluss auf die Konzentration von BRs (A), JAs (B), tZ (C), GAs (D) und IAA (E). In P28 hoch regulierte Gene sind grün, herunter regulierte rot unterlegt.

Zum GA-Haushalt gibt es ebenfalls bereits Erkenntnisse in Bezug auf die endogene Konzentration in kolumnaren Bäumen. Da diese Untersuchungen auf chemischen Eigenschaften von GAs beruhen, wurden dabei Konzentrationen polarer und apolarer GAs [211] gemessen. Polare GAs (z.B. GA₃) sind demnach in kolumnaren Bäumen schwächer repräsentiert, während apolare (z.B. GA₉) in kolumnaren wie in nicht kolumnaren in ähnlichen Mengen vertreten sind [92]. In dieser Biosynthese wurde ein einzelnes Gen als in P28 herunter reguliert identifiziert (Abbildung 4.1 D). Dies ist GA2ox, dessen Genprodukt die Menge an bioaktiven GAs durch deren Inaktivierung reguliert und damit einen direkten Einfluss auf das Wachstum hat [212]. Wie man Abbildung 4.1 D entnehmen kann, ist GA2ox aber auch maßgeblich an der Umwandlung von apolaren in polare GAs beteiligt, wobei Polarisierung nicht zwangsläufig gleichbedeutend mit Inaktivierung ist (vgl. GA₁, GA₃). Durch nachgewiesenermaßen geringere Mengen von GA2ox im SAM kolumnarer Bäume erklärt sich so die geringere Menge polarer Gibberelline, die mit entsprechenden Bioassays gemessen werden konnte. Da die Synthese der bioaktiven, polaren GAs (GA₁, GA₃) nicht beeinträchtigt ist, müsste somit ein entsprechend höherer Anteil aktiver GAs im kolumnaren SAM vorliegen. Die entsprechenden Überexpressionsmutanten zeigen aufgrund der geringeren Menge bioaktiver GAs Zwergwuchs in der Pappel [213], in Reis [214] und in *Arabidopsis* [215]. Die entsprechenden Funktionsverlustmutanten, die das Extrem der hier gezeigten Regulation darstellen, resultieren in einem hyperelongierten Phänotyp, auch als *slender* bekannt [216, 217]. Dieser ist mit dem kolumnaren Phänotyp nicht zu vergleichen, gibt aber Hinweise darauf, warum diese Wuchsform trotz extrem gestauchter Internodien die gleiche Höhe wie normalwüchsige Bäume erreicht. Mit Betrachtung der einzelnen Faktoren, die einen erwiesenen Einfluss auf das Wachstum haben, zeichnet sich immer klarer ab, dass nur die Kombination den vorliegenden Phänotyp erklären kann.

Die Betrachtung des Auxinhaushalts ist in besonderem Maße interessant, da der säulenförmige Wuchs auf eine extreme Form der Apikaldominanz hindeutet. In der Biosynthese von IAA wurden zwei in P28 herunter regulierte Gene gefunden, die für die Konjugation aktiver IAA mit Aminosäuren sorgen (Abbildung 4.1 E), was als temporärer Speicher dient [218]. Diese Amidosynthetasen (GH3.x, [219, 220]) sorgen demnach für die Inaktivierung freier IAA. Durch die schwächere Expression von GH3.1 und GH3.2 in P28 ist die Menge an freier, aktiver IAA im SAM höher, was ebenfalls frühere Messungen bestätigt [92, 105]. Weiterhin sind insgesamt drei am Transport von IAA beteiligte Gene signifikant differentiell reguliert – die Influx und Efflux Carrier AUX1 und PIN1 (hoch reguliert) sowie der membranständige

Rezeptor ABP19A (herunter reguliert). Aufgrund der asymmetrischen Lokalisation von AUX1 (apikal) und PIN1 (basal) [221] erfolgt der Transport basipetal. Bei erhöhter Expression dieser Transporter in P28 kann somit ebenfalls von einem erhöhten, basipetal gerichteten Transport freier IAA ausgegangen werden, was zu höheren Konzentrationen im Spross und zu einer Verschiebung des IAA-Gradienten führt (Abbildung 4.2). Das damit einhergehende, erhöhte Maß von Apikaldominanz [37] und apikaler Kontrolle [38, 222] drückt sich durch weniger sylleptische und proleptische Triebe und geringerer Akrotonie aus – typische Merkmale, die kolumnare Apfelbäume aufweisen [88]. Dagegen spricht einzig die große Zahl der Fruchtspieße, die aus dormanten Blattachselknospen entstehen und deren Auswachsen damit nicht dem Prinzip der Apikaldominanz folgt. Eine mögliche Erklärung liefert hierfür jedoch die erhöhte Konzentration von CK (siehe oben) in diesen Meristemen, welches IAA als Antagonist entgegenwirkt und so das Auswachsen fördert [39, 54]. Warum diese Seitentriebe größtenteils ihr Wachstum einstellen oder nur minimal wachsen, könnte mit der großen Konkurrenz zu anderen Seitentrieben und daraus resultierend mit mangelnder Nährstoffversorgung zusammenhängen [92]. Triebe, die dieser Inhibition entkommen und weiter auswachsen, zeigen für sich genommen wieder einen einheitlich kolumnaren Wuchs, wachsen nahezu parallel zur Hauptachse und verzweigen in der Regel nicht mehr. Die Regulation des Rezeptors ABP19A [223] spricht dafür, dass IAA von den Zellen vermindert perzipiert wird und es demnach zu einer verminderten Ansäuerung der Zellwand und zur verringerten Expression früher IAA-abhängiger Gene wie Expansinen, XTHs oder GH3s kommt [224]. Dies lässt sich durch die Transkriptomdaten wie gezeigt bestätigen.

Über den ABA-Haushalt lässt sich aufgrund der hier gewonnenen Resultate keine Aussage treffen, da keine in der Biosynthese von ABA involvierten, differentiell regulierten Gene gefunden wurden. Für ABA konnten zuvor ebenfalls niedrigere Konzentrationen in kolumnaren Bäumen nachgewiesen werden [92, 111].

Welche der hier diskutierten Gene in ihrer Expression tatsächlich direkt und welche durch Interaktionen mit anderen Hormonen oder Regulatoren beeinflusst werden, sollte an weiteren Geweben verifiziert werden. Zumindest von CYP735A1, ZOG1 und SQLE konnte die gezeigte Regulation mit Hilfe von qRT-PCR an *in vitro*-Kulturen bereits bestätigt werden.

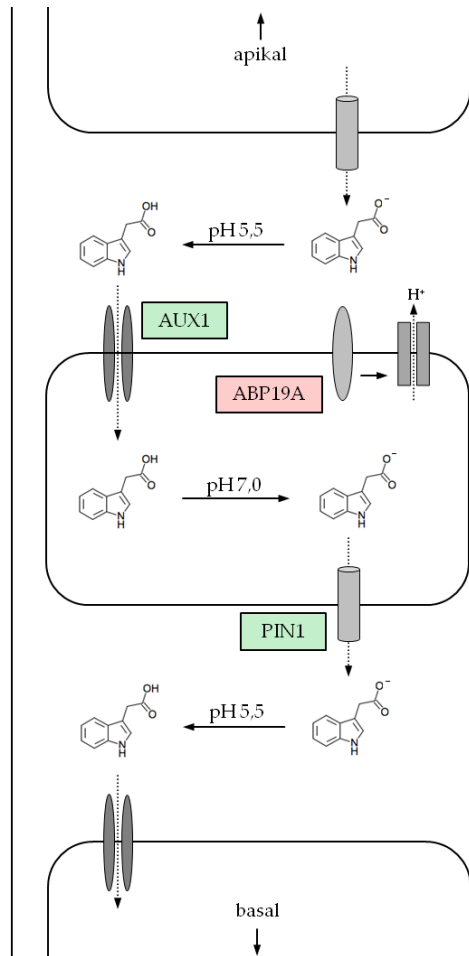


Abbildung 4.2: Polarer Transport von IAA. Der unidirektional-basipetal gerichtete Transport von IAA verläuft von Zelle zu Zelle über die Influx- und Efflux-Carrier AUX1 und PIN1, die in P28 hoch reguliert vorliegen (grün unterlegt). Diese transportieren jeweils nur die deprotonierte (Cytoplasma) bzw. protonierte Form (Interzellularraum) und sind zudem asymmetrisch lokalisiert, um den unidirektionalen Transport zu ermöglichen. IAA kann ebenfalls an Rezeptoren binden, was die Ansäuerung der Zellwand auslöst. Mit ABP19A ist ein solcher in P28 herunter reguliert (rot unterlegt).

Zusammenfassend können die Erkenntnisse der Expressionsstudien viele Charaktereigenschaften des kolumnaren Phänotyps durchaus auf molekularer Basis erklären. So überwiegen die wachstumsfördernden Phytohormone, wodurch eine normale Gesamthöhe und eine höhere Zahl von Internodien erreicht werden. Allerdings sind die Internodien im Vergleich zum normalen Habitus verkürzt, was auf Defekte in der Zellelongation, möglicherweise bedingt durch verminderte Rezeptorexpression, zurückgeführt werden kann. Durch verstärkte Synthese und Transport von aktiver IAA ist eine extrem hohe Apikaldominanz gegeben, die lediglich von der ebenfalls hohen Konzentration von CKs durchbrochen wird, was letztendlich zum Kurztriebwuchs führt. Inwiefern hierbei das Genprodukt von *Co* involviert ist, kann daraus jedoch nicht abgeleitet werden.

4.6 CO-ALLEL UND BEEINFLUSSUNG BENACHBARTER GENE

Insgesamt lassen sich 24 Gene beziffern, deren differentielle Regulation sich unabhängig vom Zeitpunkt der Entnahme mit mindestens zwei der vorgestellten Methoden nachweisen lässt (Tabelle 3.10). Zu Grunde liegen dabei die sechs in dieser Arbeit produzierten Illumina-Datensätze (Bowtie, DESeq), vier Proben verschiedener Daten während der Wachstumsperiode (Microarray) sowie *in vitro*-Kulturen (qRT-PCR). Dementsprechend basieren die gezeigten Ergebnisse auf mindestens vier (Bowtie + qRT-PCR bzw. DESeq + qRT-PCR) und maximal acht Messungen (alle Methoden) und stellen somit einen sehr hohen Grad der Signifikanz dar. Auch bei diesen 24 bestätigt sich der Trend, dass insgesamt mehr Gene in P28 hoch (15) als herunter reguliert (9) sind.

Die funktionale Charakterisierung dieser Gene liefert neben der Analyse der am Phytohormonhaushalt beteiligten und ebenfalls hier vertretenen GA2ox1, CYP735A1 und SQLE (4.5) weitere interessante Erkenntnisse in Bezug auf das Zustandekommen des kolumnaren Phänotyps. So sind die bereits mehrfach erwähnten XTHs (Tabelle 3.10 Pos. 1) maßgeblich an der Elongation von Zellen beteiligt [225]. Funktionsverlustmutanten in bereits einem Gen dieser großen Familie zeigen verkürzte Internodien aufgrund defekter Zellstreckung [226, 227], was eindeutig mit dem vorliegenden Phänotyp konform geht. Auch dem Epidermis-specific secreted glycoprotein EP1 (Pos. 3) wird eine Funktion in der Zellstreckung zugesprochen [228], wobei die Herunterregulation des Gens vermutlich ähnliche Effekte wie bei XTHs verursacht. Die Gene Thaumatin-like protein 1b [229], Bax Inhibitor 1 [230, 231] und DnaJ [232] (Pos. 2, 7, 8) haben Funktionen in der zellulären Stressantwort, und auch die Polyphenoloxidase (PPO, Pos. 5) wird normalerweise in verwundeten Geweben hoch reguliert [233]. Da alle diese Gene in P28 herunter reguliert sind, könnte dies ein Indiz dafür sein, dass kolumnare Bäume anders auf biotischen und abiotischen Stress reagieren. Ein Bezug zu möglichen, daraus resultierenden phänotypischen Auswirkungen ist jedoch wie auch bei Sieve element occlusion b [234] und Splicing factor U2af large subunit A [235] (Pos. 4, 6) nicht auszumachen.

Bei den in P28 hoch regulierten Genen ist nur vom Gen für einen Peptid/Nitrat-Transporter (Pos. 12) bekannt, dass eine Überexpression zu geänderten, in diesem Fall erhöhtem Sprosswachstum führt [236]. Alle weiteren Gene lassen sich nicht direkt mit dem Phänotyp assoziieren, allerdings tragen Vacuolar cation/proton

exchanger 3 und Probable inactive purple acid phosphatase 27 (Pos. 13, 21) zur Zellhomöostase bei, indem sie H^+/Ca^{2+} - bzw. Fe/Mn-Konzentrationen regulieren [193, 237]. Daraus resultierende Konzentrationsänderungen dieser Ionen, insbesondere Ca^{2+} , könnten wiederum in der geänderten Genexpression Ca^{2+} -abhängiger Gene resultieren. Thioredoxin O1 (Pos. 10) ist Bestandteil des mitochondrialen Thioredoxin-Systems und vermutlich am Schutz vor oxidativem Stress beteiligt [238]. Generell scheinen mitochondriale Funktionen in den Zellen kolumnarer Apfelbäume in einer bislang unbekanntem Art und Weise beeinträchtigt zu sein, was aus der Überrepräsentation des BINs ‚mitochondrialer Elektronentransport‘ (Abbildung 3.13) hervorgeht.

Um mögliche Zusammenhänge in der Regulation dieser Gene zu untersuchen, wurden die 5'-UTRs der 24 Gene auf bekannte und gemeinsame Promotorelemente hin analysiert. Allerdings ergaben sich hieraus keine überrepräsentierten Motive, wodurch ein gemeinsamer Transkriptionsfaktor weitgehend ausgeschlossen werden kann. Vielversprechend für das Auffinden eines zentralen Elements, welches derlei Regulation potentiell verursachen kann (*Co*-Gen), stellt auch die Analyse der Protein-Protein-Interaktionen mit STRING (Search Tool for the Retrieval of Interacting Genes) [239] dar. Dies liefert für die hier beschriebenen Gene allerdings keinerlei eindeutige Ergebnisse, was die Vermutung nahelegt, dass viele der Gene einem indirekten Einfluss unterliegen.

Auffällig und statistisch als signifikant belegt ist jedoch eine Anreicherung dieser differentiell regulierten Gene auf Chromosom 10, auf dem ebenfalls das *Co*-Allel lokalisiert ist (Abbildung 1.8). Vier dieser sieben Gene lassen sich darüber hinaus auf direktem Wege mit dem kolumnaren Phänotyp assoziieren: CYP735A1, SQLE, XTH und GA2ox1 (vgl. 4.5). Alle sieben Gene (und vermutlich weitere, bislang noch nicht identifizierte) könnten demnach direkt vom *Co*-Genprodukt oder damit assoziierten Faktoren in ihrer Regulation beeinflusst werden (Abbildung 4.3 B) und über ihre Genprodukte wiederum andere Gene regulieren, weshalb die übrigen 17 Gene in ihrer Lokalisation weitgehend normalverteilt vorliegen. Sollte diese Hypothese stimmen, wäre eine Regulation auf epigenetischer Ebene naheliegend, die sowohl über nicht-kodierende RNAs (RNAi, Abbildung 4.3 A) wie auch über Histonmodifikationen (Abbildung 4.3 C) realisierbar ist. Zwar liegt ein entsprechendes Gen für eine Lysin-Demethylase in der im Fokus stehenden Region, allerdings ist eine überwiegend lokale, gleichzeitig aktivierende wie reprimierende Wirkung des Genproduktes selbst mutiert schwer vorstellbar. Zudem wird für diese

Demethylase nur eine regulatorisch hemmende Funktion postuliert [240, 241]. Wahrscheinlicher ist (sofern epigenetische Regulation überhaupt eine Rolle spielt) die Regulation durch ncRNAs (non-coding RNAs), da diese wie im Falle von XIST [242] nur einen lokalen Einfluss, in diesem Fall ebenfalls auf ein gesamtes Chromosom, haben können.

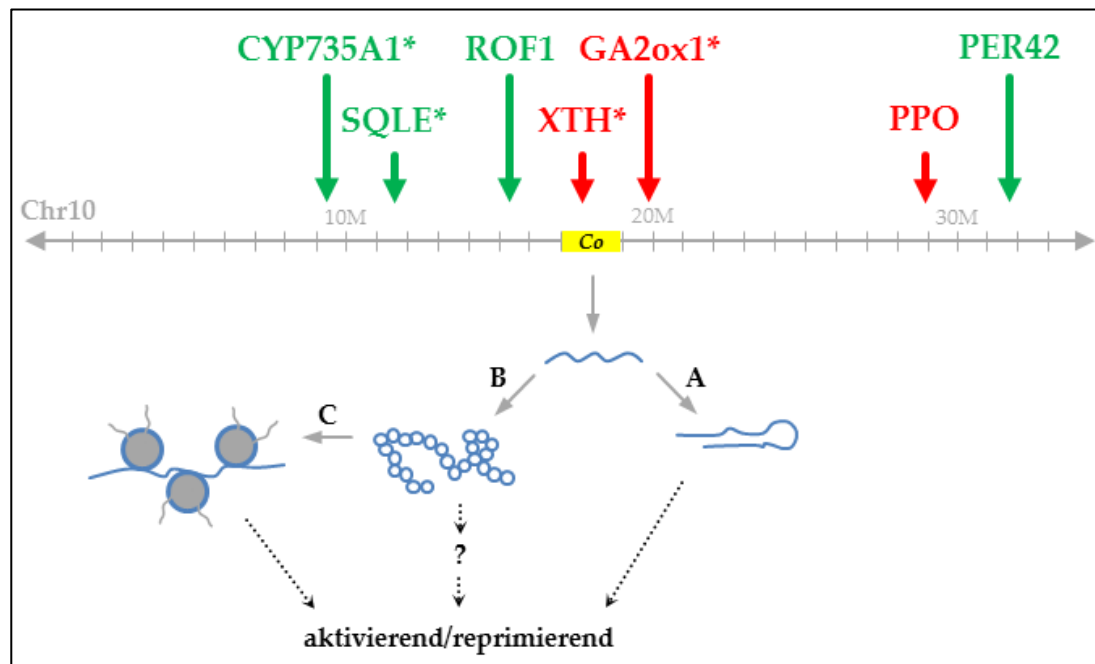


Abbildung 4.3: Mögliche Funktionen des Co-Genproduktes. Anhand der Anreicherung differenziell regulierter Gene auf Chromosom 10 (oben), von denen vier direkt assoziierbar mit dem kolumnaren Phänotyp sind (*), lassen sich drei mögliche Arten der Regulation ableiten, die sowohl aktivierend wie auch reprimierend auf die gezeigten Gene wirken. Dies können sein: **A)** ncRNAs, **B)** das Genprodukt oder damit assoziierte Faktoren (?), **C)** Histonmodifikationen, ebenfalls bedingt durch das Genprodukt von *Co*.

Eine weitere Möglichkeit besteht darin, dass der *Co*-Locus einem „selective sweep“ [243-245] unterliegt und die Regulation der gezeigten Gene Voraussetzung dafür ist, dass sich der kolumnare Phänotyp durch die vorhandene Mutation überhaupt ausdrücken kann. Dafür würde sprechen, dass bei Kreuzungsexperimenten oftmals nicht ganz 50 % kolumnare Nachkommen (durch Rekombination nahe des *Co*-Locus) auftreten und selbst kolumnare Nachkommen unterschiedliche Grade der phänotypischen Ausprägung zeigen (vgl. 1.3 und Peter Braun, persönliche Kommunikation). Ebenfalls dafür spräche der Nachweis von *Co*-gekoppelten Markern auf Chromosom 3 (Abbildung 1.8) und einem signifikant differentiell

regulierten, dort lokalisierten Gen (EP1; 28,7 Mb), was auf ein mit einem „selective sweep“ verknüpften „linkage disequilibrium“ [246] hindeutet.

Bevor diesen Hypothesen allerdings nachgegangen werden kann, sollte zunächst die Expression der besagten Gene in zusätzlichen Geweben kolumnarer Bäume untersucht werden, um eine gewebespezifische Expression auszuschließen. Darüber hinaus würde die Untersuchung der Genexpression in Sorten mit verschiedenen Ausprägungen des Kurztriebwachses bei Abwesenheit des *Co*-Allels [247, 248] weitere interessante Erkenntnisse liefern.

Aus den Ergebnissen der Expressionsstudien lassen sich insgesamt drei Gene identifizieren, die neben signifikant differentieller Regulation eine Lokalisation in der kartierten *Co*-Region aufweisen, was diese Gene zu potentiellen Kandidatengenen für *Co* macht. Dies sind das bereits mehrfach genannte XTH (17,7 Mb, herunter reguliert), eine putative Methyltransferase PMT8 (18,3 Mb) und ein NAC domain-containing protein (ANAC, 18,4 Mb), die beide in P28 hoch reguliert sind. XTH liegt jedoch nach neuesten Kartierungsergebnissen (Dominik Otto, nicht publizierte Daten) außerhalb der im Fokus stehenden Region. Für dieses Gen konnte zwar eine potentielle Beteiligung am Zustandekommen des kolumnaren Phänotyps gezeigt werden, die Herunterregulierung lässt sich jedoch schwer mit einem dominanten Effekt des *Co*-Allels in Einklang bringen. Für PMT8 ist lediglich eine subzelluläre Lokalisation in der Membran des Golgi-Apparates bekannt [249], jedoch nichts darüber, was das Substrat für die Methyltransferaseaktivität angeht. ANAC-Proteine hingegen haben eine DNA Bindedomäne und stellen in der Regel Transkriptionsaktivatoren dar. So sind sie unter anderem an der Aktivierung von JA-induzierten Genen beteiligt [250] und in Prozesse involviert, die die Formierung des SAMs regulieren [251]. Dies und die Annahme, dass durch eine mögliche Hyperaktivierung eines Transkriptionsaktivators die Regulation zahlreicher Gene beeinflusst wird, macht dieses ANAC-Protein (MDP0000214258) zu einem potentiellen Kandidaten für *Co*.

5 ZUSAMMENFASSUNG

Kolumnare Apfelbäume stellen aufgrund ihres auffälligen Phänotyps eine ökonomisch interessante Wuchsform dar. Diese extreme Form des Kurztriebwachses zeichnet sich durch einen insgesamt sehr schlanken, säulenförmigen Habitus aus, welcher eine dichte Pflanzung und damit einhergehend Ertragssteigerungen im Vergleich zu normalwüchsigen Bäumen ermöglicht. Verursacht wird der Phänotyp durch die Anwesenheit eines einzelnen, dominanten Allels des *Co*-Gens. Bis auf die approximative Lokalisation des Gens auf Chromosom 10 ist über mögliche Identität und Funktion bislang nichts bekannt.

In der vorliegenden Arbeit wurde ein erster Versuch unternommen, mit Hilfe von Next Generation Sequencing (NGS) Technologien Einblicke in das Transkriptom (RNA-Seq) des Sprossapikalmeristems (SAM) kolumnarer Apfelbäume zu gewinnen. So konnte gezeigt werden, dass unabhängig vom Zeitpunkt der Entnahme des Materials mehrere hundert Gene differentiell reguliert werden. Diese lassen sich funktional in mehrere überrepräsentierte Kategorien gruppieren, von denen sich einige wiederum mit dem kolumnaren Phänotyp assoziieren lassen. Durch den Einsatz weiterer Expressionsstudien (Microarrays, qRT-PCR) konnten frühere Ergebnisse bezüglich des Hormonhaushalts auf Genebene bestätigt und neue Erkenntnisse gewonnen werden, die eine mögliche Erklärung für den Phänotyp darstellen. Weiterhin ergab der Vergleich aller durchgeführten Expressionsstudien eine Anreicherung signifikant differentiell regulierter Gene auf Chromosom 10, weshalb *Co* eine potentielle Funktion in der epigenetischen Regulation dieser Gene einnehmen könnte. Mehr als die Hälfte dieser Gene lassen sich darüber hinaus aufgrund ihrer Funktion direkt mit dem kolumnaren Phänotyp assoziieren.

Diese Ergebnisse zeigen, dass die Anwesenheit des *Co*-Allels massive Veränderungen in der Genregulation des SAMs mit sich bringt, wobei einige dieser differentiell regulierten Gene mit großer Wahrscheinlichkeit an der Etablierung des kolumnaren Phänotyps beteiligt sind. Auch wenn die Funktion des *Co*-Genproduktes nicht abschließend geklärt werden konnte, sind doch anhand der Resultate schlüssige Hypothesen diesbezüglich möglich.

ABKÜRZUNGSVERZEICHNIS

A14	A14-190-93 (nicht-kolumnare Apfelsorte)
ABP	Auxin-binding protein
ADF	Actin Depolymerizing Factor
AHK	Histidine Kinase
AHP	<i>Arabidopsis</i> Histidine Phosphotransferase
Amp	Ampicillin
AOC4	Allene Oxide Cyclase 4
AOS	Allene Oxide Synthase
ARF	Auxin Response Factor
aRNA	antisenseRNA
ARR	<i>Arabidopsis</i> Response Regulators
AS	Aminosäure
AUX1	Auxin Transporter Protein 1
BAC	Bacterial Artificial Chromosome
bHLH	basische Helix-Loop-Helix
BI-1	Bax Inhibitor 1
BL	Brassinolid
BLAST	Basic Local Alignment Search Tool
BORT	Blast Output Refinement Tool
bp	Basenpaar(e)
BRI1	BR Insensitive 1
BRs	Brassinosteroide
βTub	Beta-Tubulin
BWT	Burrows-Wheeler-Transformation
CAX3	Vacuolar Cation/Proton Exchanger 3
cDNA	komplementäre DNA
CK	Cytokinin
cM	Centimorgan
COI1	Coronatine-insensitive protein 1
CRE1	Cytokinin-resistant
CUL	Cullin
CYP	Cytochrom P450 Monooxygenasen
CYP735A1	Cytokinin hydroxylase
DAPI	4',6-Diamidin-2-phenylindol
DIG	Digoxygenin

DIM/DWF1	Diminuto/Dwarf 1
DMAPP	Dimethylallylpyrophosphat
DMF	Dimethylformamid
ds cDNA	doppelsträngige cDNA
EDTA	Ethylendiamintetraacetat
EF1a	Elongation Factor 1 alpha
EP1	Epidermis-specific secreted glycoprotein EP1
es cDNA	einzelsträngige cDNA
EST	Estimated Sequence Tag
FA	Forschungsanstalt
FISH	Fluoreszenz <i>in situ</i> -Hybridisierung
FM	Full-text Minute-space
GA2ox1	Gibberellin 2-beta-dioxygenase 1
GAPDH	Glycerinaldehyd-3-Phosphat-Dehydrogenase
Gb	Gigabase(n)
GB	Gigabyte
GFF	General Feature Format
GH3.1	Indole-3-acetic acid-amido synthetase GH3.1
GH3.2	Indole-3-acetic acid-amido synthetase GH3.2
GWD	Genomweite Duplikation
HPC	High Performance Computing
HPLC	Hochleistungsflüssigkeitschromatographie
IAA	Indol-3-essigsäure (engl. Indole-3-acetic acid)
IAA27	Auxin-responsive protein IAA27
IPA	N6-Isopentenyladenin
IPTG	Isopropyl- β -D-thiogalactopyranosid
ivK	<i>in vitro</i> -Kultur
JA	Jasmonsäure (engl. Jasmonic Acid)
JAR1	Jasmonate Resistant 1
LAX	Like AUX1
LB-Medium	Lysogeny broth
LFY	LEAFY
LOX2	Lipoxygenase 2
mA	Milliampere
Mb	Megabase(n)
MB	Megabyte
MDP	<i>Malus x domestica</i> Protein
mRNA	messengerRNA
MS	Murashige Skoog (Medium)
MYC2	Transcription factor MYC2
NCBI	National Center for Biotechnology Information
ncRNA	non-coding RNA
NGS	Next Generation Sequencing
nm	Nanometer
nt	Nukleotid(e)

P28	Procats28 (kolumnare Apfelsorte)
<i>pasta</i>	Perl-Skript zur statistischen Auswertung von Sequenzrohdaten im FASTA Format
PIN	Pin-formed
PBS	Phosphate Buffered Saline
PCR	Polymerasekettenreaktion
PER42	Peroxidase 42
PIN1	Auxin Efflux Carrier Component 1
PPO	Polyphenol Oxidase
PST	Precomputed Score Table
qRT-PCR	quantitative Real Time PCR
RAM	Random-Access Memory
RIN	RNA-Integritätswert (engl. RNA integrity number)
RNAi	RNA-Interferenz
ROF1	70kDa peptidyl-prolyl isomerase
RPKM	Reads Per Kilobase of exon model per Million mapped reads
rpm	Umdrehungen pro Minute (engl. rounds per minute)
rRNA	Ribosomale RNA
RT	Raumtemperatur
SAM	Sprossapikalmeristem
SFF	Standard Flowgram Format
SKP1	S-phase kinase-associated protein 1
SQLE	Squalene Monooxygenase
SSC	Standard-Salz-Citrat
STRING	Search Tool for the Retrieval of Interacting Genes
tZ	trans-Zeatin
TCH4	Touch Protein 4 (xyloglucan endotransglucosylase/hydrolase 22)
TF	Transkriptionsfaktor
TIR1	Transport Inhibitor Response 1
TRIS	Tris(hydroxymethyl-)aminomethan
Trp	Tryptophan
TSA	Tyramide Signal Amplification
U	Units
U2af	Splicing Factor U2af Large Subunit A
UGT87A1	UDP-glucoxyltransferase 87A1
ü. N.	über Nacht
Vol.	Volumen
X-Gal	5-Brom-4-chlor-3-indoxyl- β -D-galactopyranosid
XTH	Xyloglucan Endotransglucosylase/hydrolase
ZOG1	Zeatin-O-Glucosyltransferase 1

ABBILDUNGSVERZEICHNIS

Abbildung 1.1: Biosynthese (A) und Signaltransduktion (B) von Auxin.....	5
Abbildung 1.2: Biosynthese (A) und Signaltransduktion (B) von Cytokininen (CK) am Beispiel von trans-Zeatin (tZ)..	7
Abbildung 1.3: Biosynthese (A) und Signaltransduktion (B) von Gibberellin (GA)..	8
Abbildung 1.4: Biosynthese (A) und Signaltransduktion (B) von Brassinosteroiden (BRs) am Beispiel von Brassinolid (BL).....	10
Abbildung 1.5: Biosynthese (A) und Signaltransduktion (B) von Jasmonaten (JAs).	12
Abbildung 1.6: Biosynthese (A) und Signaltransduktion (B) von Abscisinsäure (ABA)..	13
Abbildung 1.7: Vergleich der Wuchsformen.	15
Abbildung 1.8: Lokalisation des <i>Co</i> -Gens.....	16
Abbildung 1.9: Hormonale Situation in kolumnaren Bäumen.....	17
Abbildung 2.1: Auftropfstation.....	31
Abbildung 3.1: Probenaufbereitung.....	46
Abbildung 3.2: Roche 454-Rohdaten.....	50
Abbildung 3.3: Abdeckung von ausgewählten Gensequenzen..	52
Abbildung 3.4: Contiglängenverteilung nach NGen Assemblierung..	55
Abbildung 3.5: Contiglängenverteilung nach ABySS Assemblierung.....	56
Abbildung 3.6: Datenpipeline.....	58
Abbildung 3.7: Speziesverteilung der BLAST-Treffer.....	60
Abbildung 3.8: Kladogramm.....	61
Abbildung 3.9: Contigs versus Rohdaten.....	64
Abbildung 3.10: Korrelationskoeffizienten nach Pearson.....	65
Abbildung 3.11: Streudiagramme RPKM-normalisierter Illumina-Daten.....	67
Abbildung 3.12: Genotypisch und entnahmespez. differentiell regulierte Gene.	68
Abbildung 3.13: MapMan-Visualisierung der DESeq Analyse.....	72
Abbildung 3.14: Differentiell regulierte, TF-assoziierte Gene.....	73
Abbildung 3.15: Anreicherung differentiell regulierter Gene auf Chromosom 10....	80
Abbildung 3.16: Metaphasechromosomen.....	82
Abbildung 3.17: Nachweis von rDNA.....	83
Abbildung 4.1: Hormonale Situation im SAM kolumnarer Bäume..	95
Abbildung 4.2: Polarer Transport von IAA.....	98
Abbildung 4.3: Mögliche Funktionen des <i>Co</i> -Genproduktes.....	101

TABELLENVERZEICHNIS

Tabelle 2.1: Ausgangsmaterial.....	23
Tabelle 3.1: Übersicht aller Transkriptomsequenzierungen.....	48
Tabelle 3.2: Roche 454-Rohdaten (Statistik).....	48
Tabelle 3.3: Illumina-Rohdaten (Statistik).....	49
Tabelle 3.4: Redundanz der Illumina-Daten.....	53
Tabelle 3.5: Vergleich der <i>de novo</i> Assemblierungen.....	53
Tabelle 3.6: Differenziell regulierte Gene.	63
Tabelle 3.7: Liste der 10 meistvertretenen Transkripte..	70
Tabelle 3.8: Microarray-Auswertung.....	74
Tabelle 3.9: Differenziell regulierte, Phytohormon-assoziierte Gene..	78
Tabelle 3.10: Lokalisation von signifikant differenziell regulierten Genen im Genom von <i>Malus x domestica</i>	80

LITERATURVERZEICHNIS

1. Janick J, Cummins JN, Brown SK, Hemmat M: **Apples. In Fruit Breeding, Volume I: Tree and Tropical Fruits.** Wiley & Sons, Inc.; 1996.
2. Robinson JP: **Taxonomy of the Genus Malus Mill. (Rosaceae) with Emphasis on the Cultivated Apple, Malus domestica Borkh.** *Plant Syst Evol* 2001, **226**:35-58.
3. Korban SS, Skirvin RM: **Nomenclature of the domesticated apple.** *Hort Science* 1984, **19**:177-180.
4. Morgan J, Richards A: **The book of apples:** Ebury Press, London; 1993.
5. Forsline PL, Aldwinckle HS: **Evaluation of Malus sieversii populations for disease resistance and horticultural traits.** *Acta Hort* 2004, **663**:529-534.
6. Juniper BE, Robinson J, Harris SA, Watkins R: **Encyclopedia of Genetics:** Fitzroy Dearborn, London; 2001.
7. Juniper BE, Mabberley DJ: **The story of the apple:** Timber Press; 2006.
8. Coart E, S VANG, M DEL, Larsen AS, Roldan-Ruiz I: **Chloroplast diversity in the genus Malus: new insights into the relationship between the European wild apple (Malus sylvestris (L.) Mill.) and the domesticated apple (Malus domestica Borkh.).** *Mol Ecol* 2006, **15**(8):2171-2182.
9. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D *et al*: **The genome of the domesticated apple (Malus x domestica Borkh.).** *Nat Genet* 2010, **42**(10):833-839.
10. Gerhauser C: **Cancer chemopreventive potential of apples, apple juice, and apple components.** *Planta Med* 2008, **74**(13):1608-1624.
11. Belrose: **World Apple Report.** In. Pullman: Belrose, Inc.; 2010: 5.
12. Gharghani A, Zamani Z, Talaie A, Oraguzie NC, Fatahi R, Hajnajari H, Wiedow C, Gardiner SE: **Genetic Identity and Relationships of Iranian Apple (Malus x domestica Borkh .) Cultivars and Landraces, Wild Malus Species and Representative Old Apple Cultivars Based on Simple Sequence Repeat (SSR) Marker Analysis.** *Genet Resour Crop Ev* 2009, **56**:829-842.
13. Flachowsky H, Le Roux PM, Peil A, Patocchi A, Richter K, Hanke MV: **Application of a high-speed breeding technology to apple (Malus x domestica) based on transgenic early flowering plants and marker-assisted selection.** *New Phytol* 2011, **192**(2):364-377.

14. Flachowsky H, Szankowski I, Fischer TC, Richter K, Peil A, Hofer M, Dorschel C, Schmoock S, Gau AE, Halbwirth H *et al*: **Transgenic apple plants overexpressing the Lc gene of maize show an altered growth habit and increased resistance to apple scab and fire blight.** *Planta* 2010, **231**(3):623-635.
15. Jackson JE: **Biology of apples and pears:** Camebride University Press; 2003.
16. Giovannoni J: **Harvesting the apple genome.** *Nat Genet* 2010, **42**(10):822-823.
17. Flachowsky H, Hattasch C, Hofer M, Peil A, Hanke MV: **Overexpression of LEAFY in apple leads to a columnar phenotype with shorter internodes.** *Planta* 2009, **231**(2):251-263.
18. Malnoy M, Jin Q, Borejsza-Wysocka EE, He SY, Aldwinckle HS: **Overexpression of the apple MpNPR1 gene confers increased disease resistance in Malus x domestica.** *Mol Plant Microbe Interact* 2007, **20**(12):1568-1580.
19. Zhu LH, Li XY, Welander M: **Overexpression of the Arabidopsis gai gene in apple significantly reduces plant size.** *Plant cell reports* 2008, **27**(2):289-296.
20. Sterck L, Rombauts S, Vandepoele K, Rouze P, Van de Peer Y: **How many genes are there in plants (... and why are they there)?** *Curr Opin Plant Biol* 2007, **10**(2):199-203.
21. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP *et al*: **The genome of woodland strawberry (Fragaria vesca).** *Nat Genet*, **43**(2):109-116.
22. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN *et al*: **The genome of Theobroma cacao.** *Nat Genet* 2011, **43**(2):101-108.
23. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J *et al*: **Genome sequence of the palaeopolyploid soybean.** *Nature* 2010, **463**(7278):178-183.
24. Sitte P, Weiler EW, Kadereit JW, Bresinsky A, Körner C: **Strasburger - Lehrbuch der Botanik**, vol. 35: Spektrum; 2002.
25. Traas J, Doonan JH: **Cellular basis of shoot apical meristem development.** *International review of cytology* 2001, **208**:161-206.
26. Ruther J, Meiners T, Steidle LM: **Rich in phenomena lacking in terms - A classification of kairomones:** Chemoecology; 2002.
27. Santner A, Estelle M: **Recent advances and emerging trends in plant hormone signalling.** *Nature* 2009, **459**(7250):1071-1078.
28. Gazzarrini S, McCourt P: **Cross-talk in plant hormone signalling: what Arabidopsis mutants are telling us.** *Ann Bot* 2003, **91**(6):605-612.
29. Depuydt S, Hardtke CS: **Hormone signalling crosstalk in plant growth regulation.** *Curr Biol* 2011, **21**(9):R365-373.
30. Zhang S, Wei Y, Lu Y, Wang X: **Mechanisms of brassinosteroids interacting with multiple hormones.** *Plant Signal Behav* 2009, **4**(12).
31. Bhalerao RP, Bennett MJ: **The case for morphogens in plants.** *Nature cell biology* 2003, **5**(11):939-943.

32. Zhao Y: **Auxin biosynthesis and its role in plant development.** *Annual review of plant biology* 2010, **61**:49-64.
33. Pollmann S, Duchting P, Weiler EW: **Tryptophan-dependent indole-3-acetic acid biosynthesis by 'IAA-synthase' proceeds via indole-3-acetamide.** *Phytochemistry* 2009, **70**(4):523-531.
34. Bartel B: **Auxin Biosynthesis.** *Annu Rev Plant Physiol Plant Mol Biol* 1997, **48**:51-66.
35. Li L, Hou X, Tsuge T, Ding M, Aoyama T, Oka A, Gu H, Zhao Y, Qu LJ: **The possible action mechanisms of indole-3-acetic acid methyl ester in Arabidopsis.** *Plant cell reports* 2008, **27**(3):575-584.
36. Muday GK: **Auxins and tropisms.** *Journal of plant growth regulation* 2001, **20**(3):226-243.
37. Cline MG: **The role of hormones in apical dominance. New approaches to an old problem in plant development.** *PHYSIOLOGIA PLANTARUM* 1994, **90**:230-237.
38. Wilson BF: **Apical control of branch growth and angle in woody plants.** *Am J Bot* 2000, **87**(5):601-607.
39. Domagalska MA, Leyser O: **Signal integration in the control of shoot branching.** *Nature reviews Molecular cell biology* 2011, **12**(4):211-221.
40. Friml J: **Subcellular trafficking of PIN auxin efflux carriers in auxin transport.** *Eur J Cell Biol* 2010, **89**(2-3):231-235.
41. Petrasek J, Friml J: **Auxin transport routes in plant development.** *Development* 2009, **136**(16):2675-2688.
42. Napier RM, David KM, Perrot-Rechenmann C: **A short history of auxin-binding proteins.** *Plant molecular biology* 2002, **49**(3-4):339-348.
43. Rayle DL, Cleland RE: **The Acid Growth Theory of auxin-induced cell elongation is alive and well.** *Plant physiology* 1992, **99**(4):1271-1274.
44. Chen J-G: **Dual Auxin Signaling Pathways Control Cell Elongation and Division.** *Journal of plant growth regulation* 2001, **20**:255-264.
45. Szemenyei H, Hannon M, Long JA: **TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis.** *Science* 2008, **319**(5868):1384-1386.
46. Takei K, Yamaya T, Sakakibara H: **Arabidopsis CYP735A1 and CYP735A2 encode cytokinin hydroxylases that catalyze the biosynthesis of trans-Zeatin.** *J Biol Chem* 2004, **279**(40):41866-41872.
47. Inoue T, Higuchi M, Hashimoto Y, Seki M, Kobayashi M, Kato T, Tabata S, Shinozaki K, Kakimoto T: **Identification of CRE1 as a cytokinin receptor from Arabidopsis.** *Nature* 2001, **409**(6823):1060-1063.
48. Yamada H, Suzuki T, Terada K, Takei K, Ishikawa K, Miwa K, Yamashino T, Mizuno T: **The Arabidopsis AHK4 histidine kinase is a cytokinin-binding receptor that transduces cytokinin signals across the membrane.** *Plant Cell Physiol* 2001, **42**(9):1017-1023.
49. Yonekura-Sakakibara K, Kojima M, Yamaya T, Sakakibara H: **Molecular characterization of cytokinin-responsive histidine kinases in maize.**

- Differential ligand preferences and response to cis-zeatin.** *Plant physiology* 2004, **134**(4):1654-1661.
50. Kudo T, Kiba T, Sakakibara H: **Metabolism and long-distance translocation of cytokinins.** *J Integr Plant Biol* 2010, **52**(1):53-60.
51. Mok DW, Mok MC: **Cytokinin Metabolism and Action.** *Annu Rev Plant Physiol Plant Mol Biol* 2001, **52**:89-118.
52. Argueso CT, Ferreira FJ, Kieber JJ: **Environmental perception avenues: the interaction of cytokinin and environmental response pathways.** *Plant Cell Environ* 2009, **32**(9):1147-1160.
53. Werner T, Schmulling T: **Cytokinin action in plant development.** *Curr Opin Plant Biol* 2009, **12**(5):527-538.
54. Shimizu-Sato S, Tanaka M, Mori H: **Auxin-cytokinin interactions in the control of shoot branching.** *Plant molecular biology* 2009, **69**(4):429-435.
55. To JP, Kieber JJ: **Cytokinin signaling: two-components and more.** *Trends in plant science* 2008, **13**(2):85-92.
56. Argueso CT, Raines T, Kieber JJ: **Cytokinin signaling and transcriptional networks.** *Curr Opin Plant Biol* 2010.
57. To JP, Haberer G, Ferreira FJ, Deruere J, Mason MG, Schaller GE, Alonso JM, Ecker JR, Kieber JJ: **Type-A Arabidopsis response regulators are partially redundant negative regulators of cytokinin signaling.** *The Plant cell* 2004, **16**(3):658-671.
58. Yamaguchi S: **Gibberellin metabolism and its regulation.** *Annual review of plant biology* 2008, **59**:225-251.
59. Tudzynski B, Mihlan M, Rojas MC, Linnemannstons P, Gaskin P, Hedden P: **Characterization of the final two genes of the gibberellin biosynthesis gene cluster of *Gibberella fujikuroi*: des and P450-3 encode GA4 desaturase and the 13-hydroxylase, respectively.** *J Biol Chem* 2003, **278**(31):28635-28643.
60. Tudzynski B, Rojas MC, Gaskin P, Hedden P: **The gibberellin 20-oxidase of *Gibberella fujikuroi* is a multifunctional monooxygenase.** *J Biol Chem* 2002, **277**(24):21246-21253.
61. Thomas SG, Phillips AL, Hedden P: **Molecular cloning and functional expression of gibberellin 2- oxidases, multifunctional enzymes involved in gibberellin deactivation.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(8):4698-4703.
62. Chen F, Nonogaki H, Bradford KJ: **A gibberellin-regulated xyloglucan endotransglycosylase gene is expressed in the endosperm cap during tomato seed germination.** *J Exp Bot* 2002, **53**(367):215-223.
63. Mandava NB: **Plant Growth-Promoting Brassinosteroids.** *Annu Rev Plant Physiol Plant Mol Biol* 1988, **39**:23-52.
64. Mussig C: **Brassinosteroid-promoted growth.** *Plant Biol (Stuttg)* 2005, **7**(2):110-117.
65. Symons GM, Ross JJ, Jager CE, Reid JB: **Brassinosteroid transport.** *J Exp Bot* 2008, **59**(1):17-24.

66. Fujioka S, Yokota T: **Biosynthesis and metabolism of brassinosteroids.** *Annual review of plant biology* 2003, **54**:137-164.
67. Clouse SD: **Brassinosteroid signal transduction: from receptor kinase activation to transcriptional networks regulating plant development.** *The Plant cell* 2011, **23**(4):1219-1230.
68. Ye H, Li L, Yin Y: **Recent advances in the regulation of brassinosteroid signaling and biosynthesis pathways.** *J Integr Plant Biol* 2011, **53**(6):455-468.
69. Kim TW, Wang ZY: **Brassinosteroid signal transduction from receptor kinases to transcription factors.** *Annual review of plant biology* 2010, **61**:681-704.
70. Devoto A, Turner JG: **Jasmonate-regulated Arabidopsis stress signalling network.** *PHYSIOLOGIA PLANTARUM* 2005, **123**:161-172.
71. Fonseca S, Chico JM, Solano R: **The jasmonate pathway: the ligand, the receptor and the core signalling module.** *Curr Opin Plant Biol* 2009, **12**(5):539-547.
72. Zhang Y, Turner JG: **Wound-induced endogenous jasmonates stunt plant growth by inhibiting mitosis.** *PLoS One* 2008, **3**(11):e3699.
73. Schaller F, Schaller A, Stintzi A: **Biosynthesis and Metabolism of Jasmonates.** *Journal of plant growth regulation* 2005, **23**:179-199.
74. Staswick PE, Tiryaki I: **The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in Arabidopsis.** *The Plant cell* 2004, **16**(8):2117-2127.
75. Chini A, Fonseca S, Fernandez G, Adie B, Chico JM, Lorenzo O, Garcia-Casado G, Lopez-Vidriero I, Lozano FM, Ponce MR *et al*: **The JAZ family of repressors is the missing link in jasmonate signalling.** *Nature* 2007, **448**(7154):666-671.
76. Chung HS, Koo AJ, Gao X, Jayanty S, Thines B, Jones AD, Howe GA: **Regulation and function of Arabidopsis JASMONATE ZIM-domain genes in response to wounding and herbivory.** *Plant physiology* 2008, **146**(3):952-964.
77. Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA, Browse J: **JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling.** *Nature* 2007, **448**(7154):661-665.
78. Pauwels L, Barbero GF, Geerinck J, Tilleman S, Grunewald W, Perez AC, Chico JM, Bossche RV, Sewell J, Gil E *et al*: **NINJA connects the co-repressor TOPLESS to jasmonate signalling.** *Nature* 2010, **464**(7289):788-791.
79. Zhu JK: **Salt and drought stress signal transduction in plants.** *Annual review of plant biology* 2002, **53**:247-273.
80. Nambara E, Marion-Poll A: **Abscisic acid biosynthesis and catabolism.** *Annual review of plant biology* 2005, **56**:165-185.
81. Joshi-Saha A, Valon C, Leung J: **Abscisic acid signal off the STARTing block.** *Molecular plant* 2011, **4**(4):562-580.
82. Fisher DV: **Spur-type strains of McIntosh for high density plantings.** *British Columbia Fruit Growers' Association Quarterly Report* 1969, **14**(2):3-10.

83. Blažek J: **Segregation and general evaluation of spur type or compact growth habits in apples.** *Acta Hort* 1992, **317**:71-79.
84. Eaton GW, Lapins KO: **Identification of standard and compact apple trees by discriminate function analysis.** *J Appl Ecol* 1970, **7**:267-272.
85. Tobutt KR: **Breeding columnar apples at east malling.** *Acta Hort* 1985, **159**:63-68.
86. Kenis K, Keulemans J: **Study of Tree Architecture of Apple (*Malus × domestica* Borkh.) by QTL Analysis of Growth Traits.** *Mol Breeding* 2007, **19**:193-208.
87. Jacob HB: **Breeding experiments of apple varieties with columnar growth and low chilling requirements.** *Acta Hort* 2010, **872**(159-164).
88. De Wit I, Cook NC, Keulemans J: **Characterization of Tree Architecture in two-year-old apple seedling populations of different progenies with a common columnar gene parent.** *Acta Hort* 2004, **663**:363-368.
89. Gelvonauskienė D, Gelvonauskis B, Sasnauskas A: **Impact of Rootstock on Columnar Apple Tree Growth in a Nursery.** *Scientific Works of the Lithuanian Institute of Horticulture and Lithuanian University of Agriculture* 2006, **25**:51-56.
90. Lapins KO: **Segregation of compact growth types in certain apple seedling progenies.** *Canad J Plant Sci* 1969, **49**:765-768.
91. Lapins KO, Watkins R: **Genetics of compact growth habit.** *Report of East Malling Research Station for 1972 1973*, **136**.
92. Looney NE, Lane WD: **Spur-type growth mutants of McIntosh apple: A review of their genetics, physiology and field performance.** *Acta Hort* 1984, **146**:31-46.
93. Tian YK, Wang CH, Zhang JS, James C, Dai HY: **Mapping Co, a gene controlling the columnar phenotype of apple, with molecular markers.** *Euphytica* 2005, **145**:181-188.
94. Tian YK, Wang CH, Dai HY, Zhang JS: **Screening of a RAPD marker tightly linked to Co gene in apple and the SCAR marker conversion.** *Yi chuan xue bao = Acta genetica Sinica* 2004, **31**(9):919-925.
95. Kenis K, Keulemans J, Davey MW: **Identification and stability of QTLs for fruit quality traits in apple.** *Tree Genetics & Genomes* 2008, **4**:6247-6661.
96. Kim MY, Song KJ, Hwang J-H, Shin Y-U, Lee HJ: **Development of RAPD and SCAR markers linked to the Co gene conferring columnar growth habit in apple (*Malus pumila* Mill.).** *Journal of Horticultural Science & Biotechnology* 2003, **78**(4):512-517.
97. Moriya S, Iwanami H, Kotoda N, Takahashi S, Yamamoto T, Kazuyuki A: **Development of a Marker-assisted Selection System for Columnar Growth Habit in Apple Breeding.** *J Japan Soc Hort Sci* 2009, **78**(3):279-287.
98. Zhu YD, W. ZHANG, G.C. LI and T. WANG and T. WANG, Zhang W, Li GC, Wang T, Wang T: **Evaluation of inter-simple sequence repeat analysis for mapping the Co gene in apple (*Malus pumila* Mill.).** *Journal of Horticultural Science & Biotechnology* 2007, **82**(3):371-376.

99. Conner PJ, Brown SK, Weeden NF: **Randomly Amplified Polymorphic DNA-based Genetic Linkage Maps of Three Apple Cultivars.** *J Amer Soc Hort Sci* 1997, **122**(3):350-359.
100. Hemmat M, Weeden NF, Conner PJ, Brown SK: **A DNA Marker for Columnar Growth Habit in Apple Contains a Simple Sequence repeat.** *J Amer Soc Hort Sci* 1997, **122**(3):347-349.
101. Hemmat M, Weeden NF, Manganaris AG, Lawson DM: **Molecular marker linkage map for apple.** *The Journal of heredity* 1994, **85**(1):4-11.
102. Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C: **Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.).** *Molecular Breeding* 2002, **10**:217-241.
103. Otto D: **Molekulare Charakterisierung des Co Lokus von *Malus x domestica*.** *Dissertation.* Mainz: Johannes Gutenberg-Universität.
104. Watanabe M, Bessho H, Suzuki A, Komori S: **Seasonal Changes of IAA and Cytokinin in Shoots of Columnar Type Apple Trees.** *Acta Horti* 2008, **774**:75-80.
105. Watanabe M, Suzuki A, Komori S, Bessho H: **Comparison of Endogenous IAA and Cytokinins in Shoots of Columnar and Normal Type Apple Trees.** *J Japan Soc Hort Sci* 2004, **73**(1):19-24.
106. Watanabe M, Suzuki A, Komori S, Bessho H: **Effects of Heading-back Pruning on Shoot Growth and IAA and Cytokinin Concentrations at Bud Burst of Columnar-type Apple Trees.** *J Japan Soc Hort Sci* 2006, **75**(3):224-230.
107. Jindal KK, Dalbro S: **Endogenous growth substances in normal and dwarf mutants of Cortland and Golden Delicious apple shoots.** *Physiol Plant* 1974, **32**:71-77.
108. Martin GC, Stahly EA: **Endogenous growth regulating factors in bark of EM IX and XVI apple trees.** *Proc Amer Soc Hort Sci* 1967, **91**:31-38.
109. Lane WD, Looney NE: **A Selective Tissue Culture Medium for Growth of Compact (Dwarf) Mutants of Apple.** *Theor Appl Genet* 1982, **61**:219-223.
110. Sarwar M, Skirvin RM: **Effect of thidiazuron and 6-benzylaminopurine on adventitious shoot regeneration from leaves of three strains of 'McIntosh' apple (*Malus X domestica* Borkh.) in vitro** *Scientia Horticulturae* 1997, **86**(1-4):95-100.
111. Lee JM, Looney NE: **Absisic Acid Levels and Genetic Compaction in Apple Seedlings.** *Canad J Plant Sci* 1977, **57**(1):81-85.
112. Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM: **LEAFY controls floral meristem identity in Arabidopsis.** *Cell* 1992, **69**(5):843-859.
113. Metzker ML: **Sequencing technologies - the next generation.** *Nature reviews Genetics* 2010, **11**(1):31-46.
114. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nature reviews Genetics* 2009, **10**(1):57-63.
115. Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P: **Real-time DNA sequencing using detection of pyrophosphate release.** *Analytical biochemistry* 1996, **242**(1):84-89.

116. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B: **Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(15):8817-8822.
117. Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G: **BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies.** *Nucleic Acids Res* 2006, **34**(3):e22.
118. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proceedings of the National Academy of Sciences of the United States of America* 1977, **74**(12):5463-5467.
119. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF *et al*: **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science* 1991, **252**(5013):1651-1656.
120. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270**(5235):484-487.
121. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**(5235):467-470.
122. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: **High density synthetic oligonucleotide arrays.** *Nat Genet* 1999, **21**(1 Suppl):20-24.
123. Marguerat S, Bahler J: **RNA-seq: from technology to biology.** *Cellular and molecular life sciences : CMLS* 2010, **67**(4):569-579.
124. Shendure J: **The beginning of the end for microarrays?** *Nat Methods* 2008, **5**(7):585-587.
125. Matsumura H, Bin Nasir KH, Yoshida K, Ito A, Kahl G, Kruger DH, Terauchi R: **SuperSAGE array: the direct use of 26-base-pair transcript tags in oligonucleotide arrays.** *Nat Methods* 2006, **3**(6):469-474.
126. Matsumura H, Kruger DH, Kahl G, Terauchi R: **SuperSAGE: a modern platform for genome-wide quantitative transcript profiling.** *Current pharmaceutical biotechnology* 2008, **9**(5):368-374.
127. Dugas DV, Monaco MK, Olsen A, Klein RR, Kumari S, Ware D, Klein PE: **Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid.** *BMC Genomics* 2011, **12**:514.
128. Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR *et al*: **The developmental dynamics of the maize leaf transcriptome.** *Nat Genet* 2010, **42**(12):1060-1067.
129. Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, Sun J, Li YY, Chen Q, Xia T *et al*: **Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds.** *BMC Genomics* 2011, **12**:131.
130. Sun C, Li Y, Wu Q, Luo H, Sun Y, Song J, Lui EM, Chen S: **De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis.** *BMC Genomics* 2010, **11**:262.

131. Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L, Perrotta G: **Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development.** *BMC Genomics* 2009, **10**:399.
132. Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, Wheeler N, Sederoff R, Carlson JE: **Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection.** *BMC plant biology* 2009, **9**:51.
133. Logacheva MD, Kasianov AS, Vinogradov DV, Samigullin TH, Gelfand MS, Makeev VJ, Penin AA: **De novo sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*).** *BMC Genomics* 2011, **12**:30.
134. Murashige T, Skoog F: **A revised medium for rapid growth and bio assays with tobacco tissue cultures.** *PHYSIOLOGIA PLANTARUM* 1962, **15**:473-497.
135. Fleige S, Pfaffl MW: **RNA integrity and the effect on the real-time qRT-PCR performance.** *Mol Aspects Med* 2006, **27**(2-3):126-139.
136. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T: **The RIN: an RNA integrity number for assigning integrity values to RNA measurements.** *BMC Mol Biol* 2006, **7**:3.
137. Grant SG, Jessee J, Bloom FR, Hanahan D: **Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants.** *Proceedings of the National Academy of Sciences of the United States of America* 1990, **87**(12):4645-4649.
138. Dower WJ, Miller JF, Ragsdale CW: **High efficiency transformation of *E. coli* by high voltage electroporation.** *Nucleic Acids Res* 1988, **16**(13):6127-6145.
139. Bertani G: **Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*.** *J Bacteriol* 1951, **62**(3):293-300.
140. Martin R, Busch W, Herrmann RG, Wanner G: **Efficient preparation of plant chromosomes for high-resolution scanning electron microscopy.** *Chromosome Res* 1994, **2**(5):411-415.
141. Zhong XB, Hans de Jong J, Zabel P: **Preparation of tomato meiotic pachytene and mitotic metaphase chromosomes suitable for fluorescence in situ hybridization (FISH).** *Chromosome Res* 1996, **4**(1):24-28.
142. Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE *et al*: **De novo transcriptome assembly with ABySS.** *Bioinformatics* 2009, **25**(21):2872-2877.
143. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I: **ABySS: a parallel assembler for short read sequence data.** *Genome Res* 2009, **19**(6):1117-1123.
144. Pevzner PA, Tang H, Waterman MS: **An Eulerian path approach to DNA fragment assembly.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(17):9748-9753.

145. Zhang Y, Waterman MS: **An Eulerian path approach to global multiple alignment for DNA sequences.** *J Comput Biol* 2003, **10**(6):803-819.
146. Martinez HM: **An efficient method for finding repeats in molecular sequences.** *Nucleic Acids Res* 1983, **11**(13):4629-4634.
147. Needleman SB, Wunsch CD: **A general method applicable to the search for similarities in the amino acid sequence of two proteins.** *J Mol Biol* 1970, **48**(3):443-453.
148. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
149. Smith TF, Waterman MS: **Identification of common molecular subsequences.** *J Mol Biol* 1981, **147**(1):195-197.
150. UniProt-Consortium: **The Universal Protein Resource (UniProt) in 2010.** *Nucleic Acids Res* 2010, **38**(Database issue):D142-148.
151. Stekel DJ, Git Y, Falciani F: **The comparison of gene expression from multiple cDNA libraries.** *Genome Res* 2000, **10**(12):2055-2061.
152. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**(7):621-628.
153. Campagna D, Albiero A, Bilardi A, Caniato E, Forcato C, Manavski S, Vitulo N, Valle G: **PASS: a program to align short sequences.** *Bioinformatics* 2009, **25**(7):967-968.
154. Langmead B: **Aligning short sequencing reads with Bowtie.** *Curr Protoc Bioinformatics* 2010, **Chapter 11**:Unit 11 17.
155. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**(3):R25.
156. Ferragina P, Manzini G: **Opportunistic data structures with applications.** *Proceedings 41st Annual Symposium on Foundations of Computer Science* 2000:390-398.
157. Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M: **MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** *Plant J* 2004, **37**(6):914-939.
158. Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M: **A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize.** *Plant Cell Environ* 2009, **32**(9):1211-1229.
159. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al*: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313**(5793):1596-1604.
160. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**(10):R106.
161. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.** *Genome Res* 2008, **18**(9):1509-1517.

162. Wang L, Feng Z, Wang X, Zhang X: **DEGseq: an R package for identifying differentially expressed genes from RNA-seq data.** *Bioinformatics* 2010, **26**(1):136-138.
163. Leon J, Rojo E, Sanchez-Serrano JJ: **Wound signalling in plants.** *J Exp Bot* 2001, **52**(354):1-9.
164. Vij S, Tyagi AK: **Emerging trends in the functional genomics of the abiotic stress response in crop plants.** *Plant Biotechnol J* 2007, **5**(3):361-380.
165. Surget-Groba Y, Montoya-Burgos JI: **Optimization of de novo transcriptome assembly from next-generation sequencing data.** *Genome Res* 2010, **20**(10):1432-1440.
166. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2007, **35**(Database issue):D61-65.
167. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M *et al*: **UniProt: the Universal Protein knowledgebase.** *Nucleic Acids Res* 2004, **32**(Database issue):D115-119.
168. Mann H, Whitney D: **On a test of whether one of two random variables is stochastically larger than the other.** *Annals of mathematical Statistics* 1947, **18**:50-60.
169. Wilcoxon F: **Individual Comparisons by Ranking Methods.** *Biometrics Bulletin* 1945, **1**:80-83.
170. Tajima Y, Imamura A, Kiba T, Amano Y, Yamashino T, Mizuno T: **Comparative studies on the type-B response regulators revealing their distinctive properties in the His-to-Asp phosphorelay signal transduction of Arabidopsis thaliana.** *Plant Cell Physiol* 2004, **45**(1):28-39.
171. Curaba J, Herzog M, Vachon G: **GeBP, the first member of a new gene family in Arabidopsis, encodes a nuclear protein with DNA-binding activity and is regulated by KNAT1.** *Plant J* 2003, **33**(2):305-317.
172. Siegfried KR, Eshed Y, Baum SF, Otsuga D, Drews GN, Bowman JL: **Members of the YABBY gene family specify abaxial cell fate in Arabidopsis.** *Development* 1999, **126**(18):4117-4128.
173. Nakano T, Suzuki K, Fujimura T, Shinshi H: **Genome-wide analysis of the ERF gene family in Arabidopsis and rice.** *Plant physiology* 2006, **140**(2):411-432.
174. Lokan S: **Expressionsanalyse von am Kolumnarwachstum des Apfels beteiligten Gengruppen.** Johannes Gutenberg-Universität Mainz; 2011.
175. Martin RC, Mok MC, Mok DW: **Isolation of a cytokinin gene, ZOG1, encoding zeatin O-glucosyltransferase from Phaseolus lunatus.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(1):284-289.
176. Busch W, Martin R, Herrmann RG: **Sensitivity enhancement of fluorescence in situ hybridization on plant chromosomes.** *Chromosome Res* 1994, **2**(1):15-20.

177. Fuchs J, Houben A, Brandes A, Schubert I: **Chromosome 'painting' in plants - a feasible technique?** *Chromosoma* 1996, **104**(5):315-320.
178. Fuchs J, Schubert I: **Localization of seed protein genes on metaphase chromosomes of *Vicia faba* via fluorescence in situ hybridization.** *Chromosome Res* 1995, **3**(2):94-100.
179. Schuster M, Fuchs J, Schubert I: **Cytogenetics in fruit breeding – localization of ribosomal RNA genes on chromosomes of apple.** *Theor Appl Genet* 1997, **94**:322-324.
180. Botling J, Edlund K, Segersten U, Tahmasebpoor S, Engstrom M, Sundstrom M, Malmstrom PU, Micke P: **Impact of thawing on RNA integrity and gene expression analysis in fresh frozen tissue.** *Diagnostic molecular pathology : the American journal of surgical pathology, part B* 2009, **18**(1):44-52.
181. Opitz L, Salinas-Riester G, Grade M, Jung K, Jo P, Emons G, Ghadimi BM, Beissbarth T, Gaedcke J: **Impact of RNA degradation on gene expression profiling.** *BMC medical genomics* 2010, **3**:36.
182. Vermeulen J, De Preter K, Lefever S, Nuytens J, De Vloed F, Derveaux S, Hellemans J, Speleman F, Vandesompele J: **Measurable impact of RNA quality on gene expression results from quantitative PCR.** *Nucleic Acids Res* 2011, **39**(9):e63.
183. Feldman AL, Costouros NG, Wang E, Qian M, Marincola FM, Alexander HR, Libutti SK: **Advantages of mRNA amplification for microarray analysis.** *BioTechniques* 2002, **33**(4):906-912, 914.
184. Wang E, Panelli M, Marincola FM: **Complementary techniques: RNA amplification for gene profiling analysis.** *Advances in experimental medicine and biology* 2007, **593**:39-53.
185. Zhu B, Xu F, Baba Y: **An evaluation of linear RNA amplification in cDNA microarray gene expression analysis.** *Molecular genetics and metabolism* 2006, **87**(1):71-79.
186. Ferreira EN, Maschietto M, Silva SD, Brentani H, Carraro DM: **Evaluation of quantitative rt-PCR using nonamplified and amplified RNA.** *Diagnostic molecular pathology : the American journal of surgical pathology, part B* 2010, **19**(1):45-53.
187. Wilhelm BT, Landry JR: **RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing.** *Methods* 2009, **48**(3):249-257.
188. Morozova O, Hirst M, Marra MA: **Applications of new sequencing technologies for transcriptome analysis.** *Annu Rev Genomics Hum Genet* 2009, **10**:135-151.
189. Krost C, Petersen R, Schmidt ER: **The transcriptomes of columnar and standard type apple trees (*Malus x domestica*) - A comparative study.** *Gene* 2012.
190. Sims D, Mendes-Pereira AM, Frankum J, Burgess D, Cerone MA, Lombardelli C, Mitsopoulos C, Hakas J, Murugaesu N, Isacke CM *et al*: **High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing.** *Genome Biol* 2011, **12**(10):R104.

191. Watanabe H, Suzuki A, Komori S, Sato H: **Effect of plant growth regulators on shoot growth of a columnar type apple tree.** *Hort Res* 2003, **2**:97-100.
192. Goujon T, Minic Z, El Amrani A, Lerouxel O, Aletti E, Lapierre C, Joseleau JP, Jouanin L: **AtBXL1, a novel higher plant (*Arabidopsis thaliana*) putative beta-xylosidase gene, is involved in secondary cell wall metabolism and plant development.** *Plant J* 2003, **33**(4):677-690.
193. Cheng NH, Pittman JK, Shigaki T, Lachmansingh J, LeClere S, Lahner B, Salt DE, Hirschi KD: **Functional association of *Arabidopsis* CAX1 and CAX3 is required for normal growth and ion homeostasis.** *Plant physiology* 2005, **138**(4):2048-2060.
194. Gallastegui N, Groll M: **The 26S proteasome: assembly and function of a destructive machine.** *Trends in biochemical sciences* 2010, **35**(11):634-642.
195. Sullivan JA, Shirasu K, Deng XW: **The diverse roles of ubiquitin and the 26S proteasome in the life of plants.** *Nature reviews Genetics* 2003, **4**(12):948-958.
196. Liscum E, Reed JW: **Genetics of Aux/IAA and ARF action in plant growth and development.** *Plant molecular biology* 2002, **49**(3-4):387-400.
197. Sakai H, Honma T, Aoyama T, Sato S, Kato T, Tabata S, Oka A: **ARR1, a transcription factor for genes immediately responsive to cytokinins.** *Science* 2001, **294**(5546):1519-1521.
198. Aoyama T, Oka A: **Cytokinin signal transduction in plant cells.** *Journal of plant research* 2003, **116**(3):221-231.
199. Chevalier F, Perazza D, Laporte F, Le Henanff G, Hornitschek P, Bonneville JM, Herzog M, Vachon G: **GeBP and GeBP-like proteins are noncanonical leucine-zipper transcription factors that regulate cytokinin response in *Arabidopsis*.** *Plant physiology* 2008, **146**(3):1142-1154.
200. Hu YX, Wang YX, Liu XF, Li JY: ***Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development.** *Cell Res* 2004, **14**(1):8-15.
201. Sarojam R, Sappl PG, Goldshmidt A, Efroni I, Floyd SK, Eshed Y, Bowman JL: **Differentiating *Arabidopsis* shoots from leaves by combined YABBY activities.** *The Plant cell* 2010, **22**(7):2113-2130.
202. Takahashi T, Gasch A, Nishizawa N, Chua NH: **The DIMINUTO gene of *Arabidopsis* is involved in regulating cell elongation.** *Genes Dev* 1995, **9**(1):97-107.
203. Rasbery JM, Shan H, LeClair RJ, Norman M, Matsuda SP, Bartel B: ***Arabidopsis thaliana* squalene epoxidase 1 is essential for root and seed development.** *J Biol Chem* 2007, **282**(23):17002-17013.
204. Wang H, Nagegowda DA, Rawat R, Bouvier-Nave P, Guo D, Bach TJ, Chye ML: **Overexpression of *Brassica juncea* wild-type and mutant HMG-CoA synthase 1 in *Arabidopsis* up-regulates genes in sterol biosynthesis and enhances sterol production and stress tolerance.** *Plant Biotechnol J* 2011, **10**(1):31-42.

205. Suza WP, Staswick PE: **The role of JAR1 in Jasmonoyl-L- isoleucine production during Arabidopsis wound response.** *Planta* 2008, **227**(6):1221-1232.
206. Fonseca S, Chini A, Hamberg M, Adie B, Porzel A, Kramell R, Miersch O, Wasternack C, Solano R: **(+)-7-iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate.** *Nat Chem Biol* 2009, **5**(5):344-350.
207. Wasternack C: **Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development.** *Ann Bot* 2007, **100**(4):681-697.
208. Sarwar M, Skirvin RM, Kushad M, Norton MA: **Selecting dwarf apple (*Malus x domestica* Borkh.) trees in vitro: multiple cytokinin tolerance expressed among three strains of 'McIntosh' that differ in their growth habit under field conditions.** *Plant Cell, Tissue and Organ Culture* 1998, **54**:71-76.
209. Jones OP: **Physiological basis for the compact habit of 'Wijcik-McIntosh'.** *Annu Rpt E Malling Res Sta for 1981* 1982:159-160.
210. Pineda Rodo A, Brugiere N, Vankova R, Malbeck J, Olson JM, Haines SC, Martin RC, Habben JE, Mok DW, Mok MC: **Over-expression of a zeatin O-glucosylation gene in maize leads to growth retardation and tasselseed formation.** *J Exp Bot* 2008, **59**(10):2673-2686.
211. Atzorn R: **The immunoassay of gibberellins.** *Planta* 1983, **159**:1-6.
212. Gou J, Ma C, Kadmiel M, Gai Y, Strauss S, Jiang X, Busov V: **Tissue-specific expression of Populus C(19) GA 2-oxidases differentially regulate above- and below-ground biomass growth through control of bioactive GA concentrations.** *New Phytol* 2011.
213. Busov VB, Meilan R, Pearce DW, Ma C, Rood SB, Strauss SH: **Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature.** *Plant physiology* 2003, **132**(3):1283-1291.
214. Sakamoto T, Kobayashi M, Itoh H, Tagiri A, Kayano T, Tanaka H, Iwahori S, Matsuoka M: **Expression of a gibberellin 2-oxidase gene around the shoot apex is related to phase transition in rice.** *Plant physiology* 2001, **125**(3):1508-1516.
215. Schomburg FM, Bizzell CM, Lee DJ, Zeevaart JA, Amasino RM: **Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants.** *The Plant cell* 2003, **15**(1):151-163.
216. Lester DR, Ross JJ, Smith JJ, Elliott RC, Reid JB: **Gibberellin 2-oxidation and the SLN gene of *Pisum sativum*.** *Plant J* 1999, **19**(1):65-73.
217. Martin DN, Proebsting WM, Hedden P: **The SLENDER gene of pea encodes a gibberellin 2-oxidase.** *Plant physiology* 1999, **121**(3):775-781.
218. Ludwig-Muller J: **Auxin conjugates: their role for plant development and in the evolution of land plants.** *J Exp Bot* 2011, **62**(6):1757-1773.
219. Bottcher C, Keyzers RA, Boss PK, Davies C: **Sequestration of auxin by the indole-3-acetic acid-amido synthetase GH3-1 in grape berry (*Vitis vinifera***

- L.) and the proposed role of auxin conjugation during ripening. *J Exp Bot* 2010, **61**(13):3615-3625.
220. Staswick PE, Serban B, Rowe M, Tiryaki I, Maldonado MT, Maldonado MC, Suza W: **Characterization of an Arabidopsis enzyme family that conjugates amino acids to indole-3-acetic acid.** *The Plant cell* 2005, **17**(2):616-627.
221. Kleine-Vehn J, Dhonukshe P, Swarup R, Bennett M, Friml J: **Subcellular trafficking of the Arabidopsis auxin influx carrier AUX1 uses a novel pathway distinct from PIN1.** *The Plant cell* 2006, **18**(11):3171-3181.
222. Cline MG, Sadeski K: **Is auxin the repressor signal of branch growth in apical control.** *Am J Bot* 2002, **89**:1764-1771.
223. Ohmiya A, Tanaka Y, Kadowaki K, Hayashi T: **Cloning of genes encoding auxin-binding proteins (ABP19/20) from peach: significant peptide sequence similarity with germin-like proteins.** *Plant Cell Physiol* 1998, **39**(5):492-499.
224. Effendi Y, Rietz S, Fischer U, Scherer GF: **The heterozygous abp1/ABP1 insertional mutant has defects in functions requiring polar auxin transport and in regulation of early auxin-regulated genes.** *Plant J* 2011, **65**(2):282-294.
225. Liu Y, Liu D, Zhang H, Gao H, Guo X, Wang D, Zhang X, Zhang A: **The alpha- and beta-expansin and xyloglucan endotransglucosylase/hydrolase gene families of wheat: molecular cloning, gene expression, and EST data mining.** *Genomics* 2007, **90**(4):516-529.
226. Hyodo H, Yamakawa S, Takeda Y, Tsuduki M, Yokota A, Nishitani K, Kohchi T: **Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, XTH9, in inflorescence apices is related to cell elongation in Arabidopsis thaliana.** *Plant molecular biology* 2003, **52**(2):473-482.
227. Sasidharan R, Chinnappa CC, Staal M, Elzenga JT, Yokoyama R, Nishitani K, Voeselek LA, Pierik R: **Light quality-mediated petiole elongation in Arabidopsis during shade avoidance involves cell wall modification by xyloglucan endotransglucosylase/hydrolases.** *Plant physiology* 2010, **154**(2):978-990.
228. van Engelen FA, Hartog MV, Thomas TL, Taylor B, Sturm A, van Kammen A, de Vries SC: **The carrot secreted glycoprotein gene EP1 is expressed in the epidermis and has sequence homology to Brassica S-locus glycoproteins.** *Plant J* 1993, **4**(5):855-862.
229. Oh DH, Song KJ, Shin YU, Chung WI: **Isolation of a cDNA encoding a 31-kDa, pathogenesis-related 5/thaumatin-like (PR5/TL) protein abundantly expressed in apple fruit (Nalus domestica cv. Fuji).** *Biosci Biotechnol Biochem* 2000, **64**(2):355-362.
230. Oshima R, Yoshinaga K, Ihara-Ohori Y, Fukuda R, Ohta A, Uchimiya H, Kawai-Yamada M: **The Bax Inhibitor-1 needs a functional electron transport chain for cell death suppression.** *FEBS Lett* 2007, **581**(24):4627-4632.

231. Watanabe N, Lam E: **BAX inhibitor-1 modulates endoplasmic reticulum stress-mediated programmed cell death in Arabidopsis.** *J Biol Chem* 2008, **283**(6):3200-3210.
232. Miernyk JA: **The J-domain proteins of Arabidopsis thaliana: an unexpectedly large and diverse family of chaperones.** *Cell Stress Chaperones* 2001, **6**(3):209-218.
233. Boss PK, Gardner RC, Janssen BJ, Ross GS: **An apple polyphenol oxidase cDNA is up-regulated in wounded tissues.** *Plant molecular biology* 1995, **27**(2):429-433.
234. Ruping B, Ernst AM, Jekat SB, Nordziske S, Reineke AR, Muller B, Bornberg-Bauer E, Pruber D, Noll GA: **Molecular and phylogenetic characterization of the sieve element occlusion gene family in Fabaceae and non-Fabaceae plants.** *BMC plant biology* 2010, **10**:219.
235. Sridharan V, Heimiller J, Singh R: **Genomic mRNA profiling reveals compensatory mechanisms for the requirement of the essential splicing factor U2AF.** *Mol Cell Biol* 2011, **31**(4):652-661.
236. Komarova NY, Thor K, Gubler A, Meier S, Dietrich D, Weichert A, Suter Grottemeyer M, Tegeder M, Rentsch D: **AtPTR1 and AtPTR5 transport dipeptides in planta.** *Plant physiology* 2008, **148**(2):856-869.
237. Zhu H, Qian W, Lu X, Li D, Liu X, Liu K, Wang D: **Expression patterns of purple acid phosphatase genes in Arabidopsis organs and functional analysis of AtPAP23 predominantly transcribed in flower.** *Plant molecular biology* 2005, **59**(4):581-594.
238. Laloi C, Rayapuram N, Chartier Y, Grienemberger JM, Bonnard G, Meyer Y: **Identification and characterization of a mitochondrial thioredoxin system in plants.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(24):14144-14149.
239. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguetz P, Doerks T, Stark M, Muller J, Bork P *et al*: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.** *Nucleic Acids Res* 2011, **39**(Database issue):D561-568.
240. Jeong JH, Song HR, Ko JH, Jeong YM, Kwon YE, Seol JH, Amasino RM, Noh B, Noh YS: **Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in Arabidopsis.** *PLoS One* 2009, **4**(11):e8033.
241. Yang W, Jiang D, Jiang J, He Y: **A plant-specific histone H3 lysine 4 demethylase represses the floral transition in Arabidopsis.** *Plant J* 2010, **62**(4):663-673.
242. Weakley SM, Wang H, Yao Q, Chen C: **Expression and function of a large non-coding RNA gene XIST in human cancer.** *World journal of surgery* 2011, **35**(8):1751-1756.
243. Piazza P, Bailey CD, Cartolano M, Krieger J, Cao J, Ossowski S, Schneeberger K, He F, de Meaux J, Hall N *et al*: **Arabidopsis thaliana leaf form evolved via loss of KNOX expression in leaves in association with a selective sweep.** *Curr Biol* 2010, **20**(24):2223-2228.

244. Pokalyuk C: **The effect of recurrent mutation on the linkage disequilibrium under a selective sweep.** *Journal of mathematical biology* 2012, **64**(1-2):291-317.
245. Tian F, Stevens NM, Buckler ESt: **Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106 Suppl 1**:9979-9986.
246. McVean G: **The structure of linkage disequilibrium around a selective sweep.** *Genetics* 2007, **175**(3):1395-1406.
247. Lapins KO: **Spur type growth habit in 60 apple progenies.** *J Amer Soc Hort Sci* 1974, **99**:568-572.
248. Lapins KO, Fisher DV: **Four natural spur-type mutants of McIntosh apple.** *Canad J Plant Sci* 1974, **54**(2):359-362.
249. Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C *et al*: **Mapping the Arabidopsis organelle proteome.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(17):6518-6523.
250. Bu Q, Jiang H, Li CB, Zhai Q, Zhang J, Wu X, Sun J, Xie Q, Li C: **Role of the Arabidopsis thaliana NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses.** *Cell Res* 2008, **18**(7):756-767.
251. Aida M, Ishida T, Tasaka M: **Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes.** *Development* 1999, **126**(8):1563-1570.

SUPPORTING INFORMATION [SI]

Alle hier angegebenen und beschriebenen Dateien befinden sich im beigefügten, elektronischen Anhang.

DOKUMENTE

SI DOK_1 – AS-Alignment der potentiellen Zeatin-O-Glucosyltransferase

Das PDF enthält das multiple Sequenzalignment der potentiellen Zeatin-O-Glucosyltransferase (ZOG) aus *Malus x domestica* (MDP276057) gegen verschiedene Vertreter der UDP-glucosyltransferasen (U85A1, U87A2), cis-ZOGs aus *Zea mays* (CZOG1, CZOG2) und trans-ZOG aus *Phaseolus lunatus* (ZOG_Phalu). Gezeigt sind die farbkodierten Sekundärstrukturen (α -Helix rot, β -Faltblatt blau), erstellt wurde das Alignment mit Praline (<http://www.ibi.vu.nl/programs/pralinewww/>).

SEQUENZEN

SI FASTA_1 – Mit ABySS assemblierte Illumina-Daten von A14

Die FASTA-Datei enthält alle Sequenzen der mit ABySS assemblierten Contigs mit einer minimalen Abdeckung von 10 k-meren. Die Kopfzeilen (>) enthalten Nummerierung, Anzahl der Basen sowie Anzahle der k-mere, jeweils getrennt durch Leerzeichen.

SI FASTA_2 – Mit ABySS assemblierte Illumina-Daten von P28

Die FASTA-Datei enthält alle Sequenzen der mit ABySS assemblierten Contigs mit einer minimalen Abdeckung von 10 k-meren. Die Kopfzeilen (>) enthalten Nummerierung, Anzahl der Basen sowie Anzahle der k-mere, jeweils getrennt durch Leerzeichen.

SI FASTA_3 – Mit NGen assemblierte Roche 454-Daten von A14

Die FASTA-Datei enthält alle Sequenzen der mit NGen assemblierten Contigs mit einer minimalen Abdeckung von 2 „reads“. Die Kopfzeilen (>) enthalten den Namen, Anzahl der „reads“ sowie Anzahl der Basen.

SI FASTA_4 – Mit NGen assemblierte Roche 454-Daten von P28

Die FASTA-Datei enthält alle Sequenzen der mit NGen assemblierten Contigs mit einer minimalen Abdeckung von 2 „reads“. Die Kopfzeilen (>) enthalten den Namen, Anzahl der „reads“ sowie Anzahl der Basen.

SI FASTA_5 – Microarray-Sondensequenzen

Die FASTA-Datei beinhaltet alle Sequenzen mit zugehörigem MDP (Gen) bzw. Genbezeichnung für die auf die Microarrays gespotteten Sonden.

SKRIPTE**SI Perl-Skript_1 – *qseq2fastq***

Das Skript dient der Umwandlung von QSEQ- in FASTQ-Dateien.

SI Perl-Skript_2 – *qseq2fasta*

Das Skript dient der Umwandlung von QSEQ- in FASTA-Dateien.

SI Perl-Skript_3 – *pasta*

Das Skript dient der statistischen Auswertung von FASTA-Dateien. Ausgelesen werden können u.a. Anzahl und durchschnittliche Länge der Sequenzen, N50- und N80-Wert sowie die Basenzusammensetzung.

SI Perl-Skript_4 – *poet*

Das Skript dient der Extraktion kartierter Sequenzen aus GFF-Dateien.

SI Perl-Skript_5 – *Blaster1.2*

Der Perl-basierte Wrapper dient der Initiierung hochparalleler BLAST-Suchen.

SI Perl-Skript_6 – *Bort2.4*

Das Skript dient der benutzerdefinierten Ausgabe und Auswertung von BLAST-Ergebnissen.

SI Perl-Skript_7 – *Bowtier1.2*

Der Perl-basierte Wrapper dient der Initiierung von Kartierungen mit Bowtie.

SI Perl-Skript_8 – *MORT1.0*

Das Skript dient der benutzerdefinierten Ausgabe und Auswertung von Kartierungsergebnissen.

TABELLEN

SI Tabelle_1 – Vergleich der Assemblierungen

Gegeneinander aufgetragen sind durchschnittliche Contiglänge (Basen) und Anzahl der Contigs aller vier Assemblierungen. Die Regressionsgerade zeigt die hohe Korrelation.

SI Tabelle_2 – Auswertung repräsentativer BLAST-Suchen

Die Tabelle zeigt die Zählung der Gattungen von 500.000 BLAST-Suchen gegen die NCBI Nukleotid Datenbank (nt) und die Berechnung des relativen Anteils anhand der Gesamtzahl an Datenbankeinträgen.

SI Tabelle_3 – Liste der einzigartigen Gene mit Signifikanzmaß der differentiellen Regulation aus NGen assemblierten Roche 454-Daten

Die Tabelle zeigt die nach absteigenden R_j -Werten sortierten, einzigartigen Gene mit ID, Beschreibung (UniProtKB) und „read count“ (A14 und P28).

SI Tabelle_4 – Liste der einzigartigen Gene mit Signifikanzmaß der differentiellen Regulation aus ABySS assemblierten Illumina-Daten

Die Tabelle zeigt die nach absteigenden R_j -Werten sortierten, einzigartigen Gene mit ID, Beschreibung (UniProtKB) und k-mer Count (A14 und P28).

SI Tabelle_5 – Liste der einzigartigen Gene mit Signifikanzmaß der differentiellen Regulation aus nicht assemblierten Roche 454-Rohdaten

Die Tabelle zeigt die nach absteigenden R_j -Werten sortierten, einzigartigen Gene mit ID, Beschreibung (UniProtKB) und „read count“ (A14 und P28).

SI Tabelle_6 – Liste der einzigartigen Gene mit Signifikanzmaß der differentiellen Regulation aus nicht assemblierten Illumina-Rohdaten

Die Tabelle zeigt die nach absteigenden R_j -Werten sortierten, einzigartigen Gene mit ID, Beschreibung (UniProtKB) und „read count“ (A14 und P28).

SI Tabelle_7 – Genotypisch und zeitspezifisch differentiell regulierte Gene

Die Tabelle zeigt den Vergleich der RPKM-normalisierten „read counts“ von Datensätzen zweier Entnahmezeitpunkte (jeweils A14 und P28). Aufgeführt sind Gen ID (NCBI *Malus* Unigene Datenbank), RPKM-Werte mit Farbkodierung und MapMan Visualisierung der vier definierten Kategorien.

SI Tabelle_8 – Ergebnisse der DESeq Analyse

Die Tabelle umfasst die nach baseMean (Expressionsstärke) sortierten Ergebnisse der DESeq Analyse aller sechs Illumina-Datensätze. Die differentielle Regulation ist als \log_2 „change fold“ mit farblicher Kodierung (in P28 hoch reguliert: grün, in P28 herunter reguliert: rot) angegeben.

SI Tabelle_9 – Überrepräsentierte BINs aus der MapMan Visualisierung der DESeq Ergebnisse

Aufgeführt sind überrepräsentierte BINs und subBINs mit Anzahl der darin enthaltenen Elemente (Gene) und Wahrscheinlichkeitswert p.

SI Tabelle_10 – Microarray-Setup

Die Tabelle zeigt alle 1500 Gene, für die Sonden synthetisiert und gespottet wurden. Herkunft, ID (MDP) und bestes BLAST-Resultat (UniProtKB) sind angegeben.

SI Tabelle_11 – Auswertung der Microarrays

Gezeigt ist die Auswertung der Messungen (A14 gegen P28) der vier verschiedenen Entnahmezeitpunkte im Vergleich zu den NGS Daten. Darüber hinaus sind alle Gene gesondert aufgeführt, die zu den verschiedenen Zeitpunkten die gleiche Regulation zeigen.

CURRICULUM VITAE

Clemens Krost

Oskar-von-Miller-Straße 24
60314 Frankfurt am Main
Telefon: 069 56992742, 0177 8682305 (mobil)
ClemensKrost@gmx.de



Persönliche Angaben

Geburtsdatum: 12. März 1985
Nationalität: Deutsch
Familienstand: Verheiratet

Persönliches Profil

Dynamischer Biologe mit Fokus auf Next Generation Sequencing (NGS) Technologien, insbesondere Datenverarbeitung und Etablierung von Arbeitsabläufen für diverse wissenschaftliche Ziele; sehr gute Kenntnisse in aktuellen molekular-biologischen Methoden und betriebswirtschaftlichen Vorgängen; geschult in der Betreuung von Studenten und in der Arbeit mit Kunden; bewiesene Fähigkeiten zur Initiation und Organisation komplexer Projekte; teamfähig, motiviert

Erfahrung

Wissenschaftlicher Mitarbeiter, Institut für Molekulargenetik, JGU Mainz (seit 5/2009)

- Verantwortlich für Lehre und Betreuung von Studenten
- Etablierung und Beaufsichtigung von Diplomarbeiten

Praktikum, Max Planck Institut für Polymerforschung, Mainz (7/2007 bis 8/2007)

- Eigenständige Forschungsarbeit mit Fokus auf die *in vitro* Synthese von Acetylcholinrezeptoren von *Rattus norvegicus*
- Methoden: Immunopräzipitation, *in vitro* Expression und Translation von Membranproteinen, Western Blot, Elektrische Impedanzmessungen an oberflächenassoziierten Lipid-Doppelschichten

Ausbildung

Doktorarbeit, Institut für Molekulargenetik, JGU Mainz (seit 5/2009)

- Transkriptomanalyse des Sprossapikalmeristemgewebes (SAM) beim Apfel (*Malus x domestica*)
- Kurs zur Weiterbildung: Sicherheit in der Gentechnik

- Methoden: Isolation und Aufreinigung von Nukleinsäuren, Amplifikation von Poly(A) augereinigter mRNA, Roche 454- und Illumina-Sequenzierungen (GA_{IIx} und Hi-Seq2000), Erstellung von Microarrays, qRT-PCR, Fluoreszenz *in situ*-Hybridisierung (FISH)
- Bioinformatische Methoden: *de novo* Assemblierung (ABySS, SeqMan NGen, CLC Bio), SNP/Indel Erforschung, Batch BLAST (HPC Cluster), Etablierung von Datenpipelines (inkl. Perl-Skripte), Kartierung (PASS, Bowtie), Gene Ontology Tools (DAVID, MapMan), statistische Analysen (DESeq)

Diplomarbeit in Biologie, Institut für Molekulargenetik, JGU Mainz (09/2008 bis 04/2009)

- Entwicklung eines neuartigen, PCR basierten Verfahrens zur Identifikation von Mikrosatelliten-Loci
- Methoden: Identifikation von molekularen Markern, PCR, Southern Blot, Radioaktive Hybridisierung, Klonierungen
- Abschlussnote: 1,0

Studium der Biologie, JGU Mainz (04/2004 bis 08/2008)

- Abschlussnote: 1,0

Abitur, Willigis-Gymnasium Mainz (03/2004)

- Abschlussnote: 2,1

Sonstige Fähigkeiten

Sprachen: Englisch (fließend), Französisch (Grundkenntnisse)

EDV: Microsoft Word, Excel, PowerPoint, Adobe Photoshop, Perl (Grundkenntnisse)

Weitere Aktivitäten

Ehrenamtliche Tätigkeit, Biotechnologische Studenteninitiative (btS) e.V. (seit 10/2008)

- Erster Vorsitzender der Geschäftsstelle Mainz (06/2009 bis 06/2011)
- Organisation und Durchführung interdisziplinärer Veranstaltungen für interessierte Studenten, Mitgliederakquise, Zusammenarbeit mit Firmen

Eigener Betrieb, blm catering GbR (seit 06/2006)

- Spezialisierung auf Cocktail Catering, Mitarbeiterschulung und Beratung
- Referenzen: Edelmann GmbH (Frankfurt), Autostadt GmbH (Wolfsburg), Vepro AG (Pfungstadt), Ball des Sports (Wiesbaden)

Publikationen

Krost C, Petersen R, Schmidt ER (2012): The transcriptomes of columnar and standard type apple trees (*Malus x domestica*) – a comparative study. GENE, Epub 2012 Feb 13.

PUBLIKATIONSLISTE

INTERNATIONALE JOURNALE MIT PEER-REVIEW-GUTACHTEN

Krost C, Petersen R, Schmidt ER (2012): The transcriptomes of columnar and standard type apple trees (*Malus x domestica*) – a comparative study. *GENE*, Epub 2012 Feb 13.

BEITRÄGE ZU KONFERENZEN

Krost C, Petersen R, Brauksiepe B, Braun P, Schmidt ER (2012): Deep sequencing of the shoot apical meristem transcriptome of apple (*Malus x domestica*) – what RNA-Seq tells us about columnar growth in apple. *The second International Symposium on Biotechnology of Fruit Species (Biotechfruit 2012) in Nelson, New Zealand* (accepted)

Otto D, Krost C, Brandl R, Brauksiepe B, Eimert K, Schröder, MB, Schmidt ER, Braun P (2011): Ein kombinierter Genom- und Transkriptomansatz zur Identifizierung des Kolumnargens (*Co*) beim Apfel. *Tagung der GPZ-AG 19 (Obst, Gehölze, Reben) in Dresden* (accepted)

DANKSAGUNG

Mein herzlicher Dank geht an alle Personen, die mich direkt oder indirekt bei der Erstellung dieser Arbeit unterstützt haben.

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich, die vorliegende Dissertation selbstständig und ohne fremde Hilfe angefertigt zu haben. Weiterhin versichere ich, örtlich übernommene Ausführungen anderer Autoren und an Gedankengänge Anderer anlehrende eigene Formulierungen entsprechend gekennzeichnet und die Quellen zitiert zu haben.

(Ort, Datum)

(Clemens Krost)